POSTERIOR CONTRACTION RATES FOR THE BAYESIAN APPROACH TO LINEAR ILL-POSED INVERSE PROBLEMS

By Sergios Agapiou[*]
Stig Larsson[†]
and Andrew M. Stuart[*]

University of Warwick [*] and Chalmers University of Technology [†]

We consider a Bayesian nonparametric approach to a family of linear inverse problems in a separable Hilbert space setting with Gaussian noise. We assume Gaussian priors which are conjugate to the model and present a method of identifying the posterior using its precision operator. Working with the unbounded precision operator enables us to use partial differential equations (PDE) methodology to obtain rates of contraction of the posterior distribution to a Dirac measure centered on the true solution. Our methods assume a relatively weak relation between the prior covariance, noise covariance and forward operator.

1. Introduction. The solution of inverse problems provides a rich source of applications of the Bayesian nonparametric methodology. It encompasses a wide range of applications from partial differential equations (PDEs) [2], and there is a well-developed theory of classical, non-statistical, regularization [7]. On the other hand, the area of nonparametric Bayesian statistical estimation and in particular the problem of posterior consistency has attracted a lot of interest in recent years; see for instance [9, 22, 21, 24, 25, 10, 6]. Despite this, the formulation of many of these PDE inverse problems using the Bayesian approach is in its infancy [23]. Furthermore, the development of a theory of Bayesian posterior consistency, analogous to the theory for classical regularization, is under-developed with the primary contribution being the recent paper [14]. This recent paper provides a roadmap for what is to be expected regarding Bayesian posterior consistency, but is limited in terms of applicability by the assumption of simultaneous diagonalizability of the three linear operators required to define Bayesian inversion. Our aim in this paper is to make a significant step in the theory of Bayesian posterior consistency for linear inverse problems by developing a methodology which sidesteps the need for simultaneous diagonalizability. The central idea underlying the analysis is to work with precision operators rather than covariance operators, and thereby to enable use of powerful tools from PDE theory to facilitate the analysis.

Let $\mathcal{X}$ be a separable Hilbert space, with norm $||\cdot||$ and inner product $\langle\cdot,\cdot\rangle$, and let $\mathcal{A}\colon \mathcal{D}(\mathcal{A}) \subset \mathcal{X} \to \mathcal{X}$ be a known self-adjoint and positive-definite linear operator with bounded inverse. We consider the inverse problem to find u from y, where y

is a noisy observation of $\mathcal{A}^{-1}u$. We assume the model,

$$(1.1) \qquad y = \mathcal{A}^{-1}u + \frac{1}{\sqrt{n}}\xi,$$

where $\frac{1}{\sqrt{n}}\xi$ is an additive noise. We will be particularly interested in the small noise limit where $n \to \infty$.

A popular method in the deterministic approach to inverse problems is the generalized Tikhonov-Phillips regularization method in which u is approximated by the minimizer of a regularized least squares functional: define the Tikhonov-Phillips functional

$$(1.2) \qquad J_0(u) := \frac{1}{2}\left\|\mathcal{C}_1^{-\frac{1}{2}}(y - \mathcal{A}^{-1}u)\right\|^2 + \frac{\lambda}{2}\left\|\mathcal{C}_0^{-\frac{1}{2}}u\right\|^2,$$

where $\mathcal{C}_i \colon \mathcal{X} \to \mathcal{X}$, $i = 0,1$, are bounded, possibly compact, self-adjoint positive-definite linear operators. The parameter λ is called the regularization parameter, and in the classical non-probabilistic approach the general practice is to choose it as an appropriate function of the noise size $n^{-\frac{1}{2}}$, which shrinks to zero as $n \to \infty$, in order to recover the unknown parameter u [7].

In this paper we adopt a Bayesian approach for the solution of problem (1.1), which will be linked to the minimization of J_0 via the posterior mean. We assume that the prior distribution is Gaussian, $u \sim \mu_0 = \mathcal{N}(0, \tau^2 \mathcal{C}_0)$, where $\tau > 0$ and $\mathcal{C}_0$ is a self-adjoint, positive-definite, trace class, linear operator on $\mathcal{X}$. We also assume that the noise is Gaussian, $\xi \sim \mathcal{N}(0, \mathcal{C}_1)$, where $\mathcal{C}_1$ is a self-adjoint positive-definite, bounded, but not necessarily trace class, linear operator; this allows us to include the case of white observational noise. We assume that the, generally unbounded, operators $\mathcal{C}_0^{-1}$ and $\mathcal{C}_1^{-1}$, have been maximally extended to self-adjoint positive-definite operators on appropriate domains. The unknown parameter and the noise are considered to be independent, thus the conditional distribution of the observation given the unknown parameter u is also Gaussian with distribution $y|u \sim \mathcal{N}(\mathcal{A}^{-1}u, \frac{1}{n}\mathcal{C}_1)$.

Define $\lambda = \frac{1}{n\tau^2}$ and let

$$(1.3) \qquad J(u) = nJ_0(u) = \frac{n}{2}\left\|\mathcal{C}_1^{-\frac{1}{2}}(y - \mathcal{A}^{-1}u)\right\|^2 + \frac{1}{2\tau^2}\left\|\mathcal{C}_0^{-\frac{1}{2}}u\right\|^2.$$

In finite dimensions the probability density of the posterior distribution with respect to the Lebesgue measure is proportional to $\exp(-J(u))$. This suggests that, in the infinite-dimensional setting, the posterior is Gaussian $\mu^y = \mathcal{N}(m, \mathcal{C})$, where we can identify the posterior covariance and mean by the equations

$$(1.4) \qquad \mathcal{C}^{-1} = n\mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1} + \frac{1}{\tau^2}\mathcal{C}_0^{-1}$$

and

$$(1.5) \qquad \frac{1}{n}\mathcal{C}^{-1}m = \mathcal{A}^{-1}\mathcal{C}_1^{-1}y,$$

obtained by completing the square. We present a method of justifying these expressions in Section 4. We define

$$(1.6) \qquad \mathcal{B}_\lambda = \frac{1}{n}\mathcal{C}^{-1} = \mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1} + \lambda\mathcal{C}_0^{-1}$$

and observe that the dependence of $\mathcal{B}_\lambda$ on n and τ is only through λ. Since

$$(1.7) \qquad \mathcal{B}_\lambda m = \mathcal{A}^{-1}\mathcal{C}_1^{-1}y,$$

the posterior mean also depends only on λ: $m = m_\lambda$. This is not the case for the posterior covariance $\mathcal{C}$, since it depends on n and τ separately: $\mathcal{C} = \mathcal{C}_{\lambda,n}$. In the following, we suppress the dependence of the posterior covariance on λ and n and we denote it by $\mathcal{C}$.

Observe that the posterior mean is the minimizer of the functional J, hence also of J_0, that is, the posterior mean is the Tikhonov-Phillips regularized approximate solution of problem (1.1), for the functional J_0.

In [18] and [16], formulae for the posterior covariance and mean are identified in the infinite-dimensional setting, which avoid using any of the inverses of the prior, posterior or noise covariance operators. They obtain

$$(1.8) \qquad \mathcal{C} = \tau^2\mathcal{C}_0 - \tau^2\mathcal{C}_0\mathcal{A}^{-1}(\mathcal{A}^{-1}\mathcal{C}_0\mathcal{A}^{-1} + \lambda\mathcal{C}_1)^{-1}\mathcal{A}^{-1}\mathcal{C}_0$$

and

$$(1.9) \qquad m = \mathcal{C}_0\mathcal{A}^{-1}(\mathcal{A}^{-1}\mathcal{C}_0\mathcal{A}^{-1} + \lambda\mathcal{C}_1)^{-1}y,$$

which are consistent with formulae (1.4) and (1.7) for the finite-dimensional case. In [18] this is done only for $\mathcal{C}_1$ of trace class while in [16] the case of white observational noise was included. We will work in an infinite-dimensional setting where the formulae (1.4), (1.7) for the posterior covariance and mean can be justified. Working with the unbounded operator $\mathcal{B}_\lambda$ opens the possibility of using tools of analysis, and also numerical analysis, familiar from the theory of partial differential equations.

In our analysis we always assume that $\mathcal{C}_0^{-1}$ is regularizing, that is, we assume that $\mathcal{C}_0^{-1}$ dominates $\mathcal{B}_\lambda$ in the sense that it induces stronger norms than $\mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1}$. This is a reasonable assumption since otherwise we would have $\mathcal{B}_\lambda \simeq \mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1}$ (here $\simeq$ is used loosely to indicate two operators which induce equivalent norms; we will make this notion precise in due course). This would imply that the posterior mean is $m \simeq \mathcal{A}y$, meaning that we attempt to invert the data by applying the, generally discontinuous, operator $\mathcal{A}$ [7, Proposition 2.7].

We study the consistency of the posterior μ^y in the frequentist setting. To this end, we consider data $y = y^\dagger$ which is a realization of

$$(1.10) \qquad y^\dagger = \mathcal{A}^{-1}u^\dagger + \frac{1}{\sqrt{n}}\xi, \quad \xi \sim \mathcal{N}(0, \mathcal{C}_1),$$

where $u^\dagger$ is a fixed element of $\mathcal{X}$; that is, we consider observations which are perturbations of the image of a fixed true solution $u^\dagger$ by an additive noise ξ, scaled

by $\frac{1}{\sqrt{n}}$. Since the posterior depends through its mean on the data and also through its covariance operator on the scaling of the noise and the prior, this choice of data model gives as posterior distribution the Gaussian measure $\mu_{\lambda,n}^{y^\dagger} = \mathcal{N}(m_\lambda^\dagger, \mathcal{C})$, where $\mathcal{C}$ is given by (1.4) and

$$(1.11) \qquad \mathcal{B}_\lambda m_\lambda^\dagger = \mathcal{A}^{-1} \mathcal{C}_1^{-1} y^\dagger.$$

We study the behavior of the posterior $\mu_{\lambda,n}^{y^\dagger}$ as the noise disappears ($n \to \infty$). Our aim is to show that it contracts to a Dirac measure centered on the fixed true solution $u^\dagger$. In particular, we aim to determine ε_n such that

$$(1.12) \qquad \mathbb{E}^{y^\dagger} \mu_{\lambda,n}^{y^\dagger} \left\{ u : \left\| u - u^\dagger \right\| \geq M_n \varepsilon_n \right\} \to 0, \quad \forall M_n \to \infty,$$

where the expectation is with respect to the random variable $y^\dagger$ distributed according to the data likelihood $\mathcal{N}(\mathcal{A}^{-1} u^\dagger, \frac{1}{n} \mathcal{C}_1)$.

As in the deterministic theory of inverse problems, in order to get convergence in the small noise limit, we let the regularization disappear in a carefully chosen way, that is, we will choose $\lambda = \lambda(n)$ such that $\lambda \to 0$ as $n \to \infty$. The assumption that $\mathcal{C}_0^{-1}$ dominates $\mathcal{B}_\lambda$, shows that $\mathcal{B}_\lambda$ is a singularly perturbed unbounded (usually differential) operator, with an inverse which blows-up in the limit $\lambda \to 0$. This together with equation (1.7), opens up the possibility of using the analysis of such singular limits to study posterior contraction: on the one hand, as $\lambda \to 0$, $\mathcal{B}_\lambda^{-1}$ becomes unbounded; on the other hand, as $n \to \infty$, we have more accurate data, suggesting that for the appropriate choice of $\lambda = \lambda(n)$ we can get $m_\lambda^\dagger \simeq u^\dagger$. In particular, we will choose τ as a function of the scaling of the noise, $\tau = \tau(n)$, under the restriction that the induced choice of $\lambda = \lambda(n) = \frac{1}{n\tau(n)^2}$, is such that $\lambda \to 0$ as $n \to \infty$. The last choice will be made in a way which optimizes the rate of posterior contraction ε_n, defined in (1.12). In general there are three possible asymptotic behaviors of the scaling of the prior τ^2 as $n \to \infty$, [25, 14]:

   i) $\tau^2 \to \infty$; we increase the prior spread, if we know that draws from the prior are more regular than $u^\dagger$;
   ii) τ^2 fixed; draws from the prior have the same regularity as $u^\dagger$;
   iii) $\tau^2 \to 0$ at a rate slower than $\frac{1}{n}$; we shrink the prior spread, when we know that draws from the prior are less regular than $u^\dagger$.

The problem of posterior contraction in this context is also investigated in [14] and [8]. In [14], sharp convergence rates are obtained in the case where $\mathcal{C}_0, \mathcal{C}_1$ and $\mathcal{A}^{-1}$ are simultaneously diagonalizable, with eigenvalues decaying algebraically, and in particular $\mathcal{C}_1 = I$, that is, the data are polluted by white noise. In this paper we relax the assumptions on the relations between the operators $\mathcal{C}_0, \mathcal{C}_1$ and $\mathcal{A}^{-1}$, by assuming that appropriate powers of them induce comparable norms (see Section 2). In [8], the non-diagonal case is also examined; the three operators involved are related through domain inclusion assumptions. The assumptions made in [8] can be quite restrictive in practice; our assumptions include settings not covered in [8], and in particular the case of white observational noise.

Provided the problem is sufficiently ill-posed and the true solution $u^\dagger$ is sufficiently regular, in Corollary 6.5 below, we get the convergence in (1.12) for

$$\varepsilon_n = n^{-\frac{\gamma \wedge (\Delta+1)}{2(\Delta + \gamma \wedge (\Delta+1) - 1 + s_0 + \varepsilon)}}.$$

The parameters γ, s_0 and Δ measure the regularity of the true solution, the regularity of the prior and the ill-posedness of the problem respectively, as explained in Sections 2 and 6. Even though we work in a more difficult non-diagonal setting, we get rates which, up to $\varepsilon > 0$ arbitrarily small, agree with the sharp convergence rates obtained in the diagonal case in [14] for a wide range of values of the parameter γ (Figure 1). Our rates fail to be optimal when the true solution is too regular, in particular our rates saturate earlier, and we also require more regularity from the true solution in order to get convergence. Both discrepancies are attributed to the fact that our method relies on interpolating between rates in a strong and a weak norm of the error $e = m_\lambda^\dagger - u^\dagger$; on the one hand the rate of the error in the weak norm saturates earlier and on the other hand the error in the strong norm requires additional regularity in order to converge.
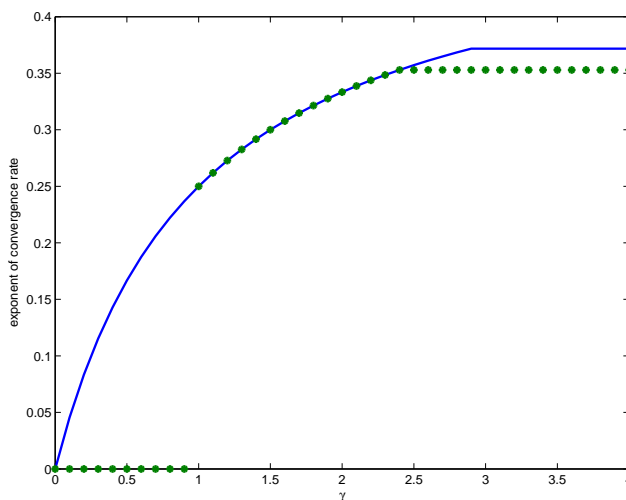

FIG 1. *Exponents of rates of contraction plotted against the regularity of the true solution, γ. In blue are the sharp convergence rates obtained in the diagonal case in [14], while in green the rates predicted by our method, which applies to the more general non-diagonal case*

In the following section we present our assumptions and their implications. In Section 3, we reformulate equation (1.7) as a weak equation in an infinite-dimensional space. In Section 4, we present a new method of identifying the posterior distribution: we first characterize it through its Radon-Nikodym derivative with respect to the prior (Theorem 4.1) and then justify the formulae (1.4), (1.7) for the posterior covariance and mean (Theorem 4.2). In Section 5, we present operator norm bounds for $\mathcal{B}_\lambda^{-1}$ in terms of the singular parameter λ, which are the key to the posterior contraction results contained in Section 6 (Theorems 6.1 and 6.2 and

Corollaries 6.4 and 6.5). In Section 7, we present a nontrivial example satisfying our assumptions and provide the corresponding rates of convergence. In Section 8, we compare our results to known minimax rates of convergence in the case where $\mathcal{C}_0, \mathcal{C}_1$ and $\mathcal{A}^{-1}$ are all diagonalizable in the same eigenbasis and have eigenvalues that decay algebraically. Finally, Section 9 is a short conclusion.

2. The Setting.   In this section we present the setting in which we formulate our results. First, we define the spaces in which we work, in particular, we define the Hilbert scale induced by the prior covariance operator $\mathcal{C}_0$. Then we determine the spaces to which draws from the prior, μ_0, and white noise belong. Furthermore, we state our main assumptions, which concern the connections between the operators $\mathcal{C}_0$, $\mathcal{C}_1$ and $\mathcal{A}^{-1}$ and present regularity results for draws from the prior, μ_0, and the noise distribution, $\mathcal{N}(0, \mathcal{C}_1)$. Finally we briefly overview the way in which the Hilbert scale defined in terms of the prior covariance operator $\mathcal{C}_0$, which is natural for our analysis, links to scales of spaces defined independently of any prior model.

2.1. *Assumptions.*   We start by defining the Hilbert scale which we will use in our analysis. Recall that $\mathcal{X}$ is an infinite-dimensional separable Hilbert space and $\mathcal{C}_0 \colon \mathcal{X} \to \mathcal{X}$ is a self-adjoint, positive-definite, trace class, linear operator. Since $\mathcal{C}_0 \colon \mathcal{X} \to \mathcal{X}$ is injective and self-adjoint we have that $\mathcal{X} = \overline{\mathcal{R}(\mathcal{C}_0)} \oplus \mathcal{R}(\mathcal{C}_0)^\perp = \overline{\mathcal{R}(\mathcal{C}_0)}$. This means that $\mathcal{C}_0^{-1} \colon \mathcal{R}(\mathcal{C}_0) \to \mathcal{X}$ is a densely defined, unbounded, symmetric, positive-definite, linear operator in $\mathcal{X}$. Hence it can be extended to a self-adjoint operator with domain $\mathcal{D}(\mathcal{C}_0^{-1}) := \{u \in \mathcal{X} : \mathcal{C}_0^{-1}u \in \mathcal{X}\}$; this is the Friedrichs extension [15]. Thus, we can define the Hilbert scale $(X^t)_{t \in \mathbb{R}}$, with $X^t := \overline{\mathcal{M}}^{\|\cdot\|_t}$ [7], where

$$\mathcal{M} := \bigcap_{k=0}^{\infty} \mathcal{D}(\mathcal{C}_0^{-k}), \ \langle u, v \rangle_t := \langle \mathcal{C}_0^{-\frac{t}{2}} u, \mathcal{C}_0^{-\frac{t}{2}} v \rangle \quad \text{and} \quad \|u\|_t := \big\| \mathcal{C}_0^{-\frac{t}{2}} u \big\|.$$

The bounded linear operator $\mathcal{C}_1 \colon \mathcal{X} \to \mathcal{X}$ is assumed to be self-adjoint, positive-definite (but not necessarily trace class); thus $\mathcal{C}_1^{-1} \colon \mathcal{R}(\mathcal{C}_1) \to \mathcal{X}$ can be extended in the same way to a self-adjoint operator with domain $\mathcal{D}(\mathcal{C}_1^{-1}) := \{u \in \mathcal{X} : \mathcal{C}_1^{-1} u \in \mathcal{X}\}$. Finally, recall that we assume that $\mathcal{A} \colon \mathcal{D}(\mathcal{A}) \to \mathcal{X}$ is a self-adjoint and positive-definite, linear operator with bounded inverse, $\mathcal{A}^{-1} \colon \mathcal{X} \to \mathcal{X}$.

Let $\{\lambda_k^2, \phi_k\}_{k=1}^{\infty}$ be orthonormal eigenpairs of $\mathcal{C}_0$ in $\mathcal{X}$. Thus, $\{\lambda_k\}_{k=1}^{\infty}$ are the singular values and $\{\phi_k\}_{k=1}^{\infty}$ an orthonormal eigenbasis. Since $\mathcal{C}_0$ is trace class we have that $\sum_{k=1}^{\infty} \lambda_k^2 < \infty$. In fact we make the following stronger assumption:

ASSUMPTION 2.1.   *There is a $\sigma_0 \in (0, 1]$ such that $\sum_{k=1}^{\infty} \lambda_k^{2(1-\sigma)} < \infty$ for all $\sigma < \sigma_0$.*

We assume that we have a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. The expected value is denoted by $\mathbb{E}$ and $\xi \sim \mu$ means that the law of the random variable ξ is the measure μ.

Let $\mu_0 := \mathcal{N}(0, \tau^2 \mathcal{C}_0)$ and $\mathbb{P}_0 := \mathcal{N}(0, \frac{1}{n} \mathcal{C}_1)$ be the prior and noise distributions respectively. Furthermore, let $\nu(du, dy)$ denote the measure constructed by taking u

and $y|u$ as independent Gaussian random variables $\mathcal{N}(0, \tau^2 \mathcal{C}_0)$ and $\mathcal{N}(\mathcal{A}^{-1}u, \frac{1}{n}\mathcal{C}_1)$ respectively:

$$\nu(du, dy) = \mathbb{P}(dy|u)\mu_0(du),$$

where $\mathbb{P} := \mathcal{N}(\mathcal{A}^{-1}u, \frac{1}{n}\mathcal{C}_1)$. We denote by $\nu_0(du, dy)$ the measure constructed by taking u and y as independent Gaussian random variables $\mathcal{N}(0, \tau^2\mathcal{C}_0)$ and $\mathcal{N}(0, \frac{1}{n}\mathcal{C}_1)$ respectively:

$$\nu_0(du, dy) = \mathbb{P}_0(dy) \otimes \mu_0(du).$$

In the following, we exploit the regularity properties of a white noise to determine the regularity of draws from the prior and the noise distributions. We consider a white noise to be a draw from $\mathcal{N}(0, I)$, that is a random variable $\zeta \sim \mathcal{N}(0, I)$. Even though the identity operator is not trace class in $\mathcal{X}$, it is trace class in a bigger space X^{-s}, where $s > 0$ is sufficiently large.

LEMMA 2.2. *Under the Assumption 2.1 we have:*

i) *Let ζ be a white noise. Then $\mathbb{E}\big\|\mathcal{C}_0^{\frac{s}{2}}\zeta\big\|^2 < \infty$ for all $s > s_0 := 1 - \sigma_0$.*
ii) *Let $u \sim \mu_0$. Then $u \in X^\sigma$ μ_0-a.s. for every $\sigma < \sigma_0$.*

PROOF.

i) We have that $\mathcal{C}_0^{\frac{s}{2}}\zeta \sim \mathcal{N}(0, \mathcal{C}_0^s)$, thus $\mathbb{E}\big\|\mathcal{C}_0^{\frac{s}{2}}\zeta\big\|^2 < \infty$ is equivalent to $\mathcal{C}_0^s$ being of trace class. By the Assumption 2.1 it suffices to have $s > 1 - \sigma_0$.
ii) We have $\mathbb{E}\big\|\mathcal{C}_0^{-\frac{\sigma}{2}}u\big\|^2 = \mathbb{E}\big\|\mathcal{C}_0^{\frac{1-\sigma}{2}}\mathcal{C}_0^{-\frac{1}{2}}u\big\|^2 = \mathbb{E}\big\|\mathcal{C}_0^{\frac{1-\sigma}{2}}\zeta\big\|^2$, where ζ is a white noise, therefore using part (i) we get the result.

$\square$

REMARK 2.3. *Note that as σ_0 changes, both the Hilbert scale and the decay of the coefficients of a draw from μ_0 change. The norms $\|\cdot\|_t$ are defined through powers of the eigenvalues λ_k^2. If $\sigma_0 < 1$, then $\mathcal{C}_0$ has eigenvalues that decay like $k^{-\frac{1}{s_0}}$, thus an element $u \in X^t$ has coefficients $\langle u, \phi_k \rangle$, that decay faster than $k^{-\frac{1}{2} - \frac{t}{2s_0}}$. As σ_0 gets larger, that is, as s_0 gets closer to zero, the space X^t for a fixed $t > 0$, corresponds to a faster decay rate of the coefficients. At the same time, by the last lemma, draws from $\mu_0 = \mathcal{N}(0, \mathcal{C}_0)$ belong to X^σ for all $\sigma < \sigma_0$. Consequently, as σ_0 gets larger, not only do draws from μ_0 belong to X^σ for larger σ, but also the spaces X^σ for fixed σ reflect faster decay rates of the coefficients. The case $\sigma_0 = 1$ corresponds to $\mathcal{C}_0$ having eigenvalues that decay faster than any negative power of k. A draw from μ_0 in that case has coefficients that decay faster than any negative power of k.*

We now state a number of assumptions regarding interrelations between the three operators $\mathcal{C}_0$, $\mathcal{C}_1$ and $\mathcal{A}^{-1}$; these assumptions reflect the idea that

$$\mathcal{C}_1 \simeq \mathcal{C}_0^\beta \quad \text{and} \quad \mathcal{A}^{-1} \simeq \mathcal{C}_0^\ell,$$

for some $\beta \geq 0, \ell \geq 0$, where $\simeq$ is used in the same manner as in Section 1. This is made precise by the inequalities presented in the following assumption, where the notation $a \asymp b$ means that there exist constants $c, c' > 0$ such that $ca \leq b \leq c'a$.

ASSUMPTION 2.4. *Suppose there exist* $\beta \geq 0$, $\ell \geq 0$ *and constants* $c_i > 0, i = 1, .., 4$ *such that, for* $\Delta := 2\ell - \beta + 1$, *we have*

1. $\Delta > 2s_0$;
2. $\left\| \mathcal{C}_1^{-\frac{1}{2}} \mathcal{A}^{-1} u \right\| \asymp \left\| \mathcal{C}_0^{\ell - \frac{\beta}{2}} u \right\|, \quad \forall u \in X^{\beta - 2\ell}$;
3. $\left\| \mathcal{C}_0^{-\frac{\rho}{2}} \mathcal{C}_1^{\frac{1}{2}} u \right\| \leq c_1 \left\| \mathcal{C}_0^{\frac{\beta - \rho}{2}} u \right\|, \quad \forall u \in X^{\rho - \beta}, \forall \rho < \beta - s_0$;
4. $\left\| \mathcal{C}_0^{\frac{s}{2}} \mathcal{C}_1^{-\frac{1}{2}} u \right\| \leq c_2 \left\| \mathcal{C}_0^{\frac{s - \beta}{2}} u \right\|, \quad \forall u \in X^{\beta - s}, \forall s \in (s_0, 1]$;
5. $\left\| \mathcal{C}_0^{-\frac{s}{2}} \mathcal{C}_1^{-\frac{1}{2}} \mathcal{A}^{-1} u \right\| \leq c_3 \left\| \mathcal{C}_0^{\frac{2\ell - \beta - s}{2}} u \right\|, \quad \forall u \in X^{s + \beta - 2\ell}, \forall s \in (s_0, 1]$;
6. $\left\| \mathcal{C}_0^{\frac{\eta}{2}} \mathcal{A}^{-1} \mathcal{C}_1^{-1} u \right\| \leq c_4 \left\| \mathcal{C}_0^{\frac{\eta}{2} + \ell - \beta} u \right\|, \quad \forall u \in X^{2\beta - 2\ell - \eta}, \forall \eta \in [\beta - 2\ell, 1]$;

where $s_0 = 1 - \sigma_0 \in [0, 1)$ *is given by Assumption 2.1.*

Notice that, by Assumption 2.4(1) we have $2\ell - \beta > -1$ which, in combination with Assumption 2.4(2), implies that

$$\left\langle \mathcal{C}_1^{-\frac{1}{2}} \mathcal{A}^{-1} u, \mathcal{C}_1^{-\frac{1}{2}} \mathcal{A}^{-1} u \right\rangle + \lambda \left\langle \mathcal{C}_0^{-\frac{1}{2}} u, \mathcal{C}_0^{-\frac{1}{2}} u \right\rangle \leq c \left\langle \mathcal{C}_0^{-\frac{1}{2}} u, \mathcal{C}_0^{-\frac{1}{2}} u \right\rangle, \quad \forall u \in X^1,$$

capturing the idea that the regularization through $\mathcal{C}_0$ is indeed a regularization. In fact the assumption $\Delta > 2s_0$ connects the ill-posedness of the problem to the regularity of the prior. The value of $2s_0$ becomes larger when $\mathcal{C}_0$ is less regular in which case we require a bigger value of Δ, which means a more ill-posed problem.

LEMMA 2.5. *Under the Assumptions 2.1 and 2.4 we have:*

i) $u \in X^{s_0 + \beta - 2\ell + \varepsilon}$ μ_0-*a.s. for all* $0 < \varepsilon < (\Delta - 2s_0) \wedge \sigma_0$;
ii) $\mathcal{A}^{-1} u \in \mathcal{D}(\mathcal{C}_1^{-\frac{1}{2}})$ μ_0-*a.s.*;
iii) $\xi \in X^\rho$ $\mathbb{P}_0$-*a.s. for all* $\rho < \beta - s_0$;
iv) $y \in X^\rho$ ν-*a.s. for all* $\rho < \beta - s_0$.

PROOF.

i) We can choose an ε as in the statement by the Assumption 2.4(1). By Lemma 2.2(ii), it suffices to show that $s_0 + \beta - 2\ell + \varepsilon < \sigma_0$. Indeed, $s_0 + \beta - 2\ell + \varepsilon = s_0 + 1 - \Delta + \varepsilon < 1 - s_0 = \sigma_0$.
ii) Under Assumption 2.4(2) it suffices to show that $u \in X^{\beta - 2\ell}$. Indeed, by Lemma 2.2(ii), we need to show that $\beta - 2\ell < \sigma_0$, which is true since $s_0 \in [0, 1)$ and we assume $\Delta > 2s_0 \geq s_0$, thus $2\ell - \beta + 1 > s_0$.
iii) Noting that $\zeta = \mathcal{C}_1^{-\frac{1}{2}} \xi$ is a white noise, using Assumption 2.4(3), we have by Lemma 2.2(i)

$$\mathbb{E} \|\xi\|_\rho^2 = \mathbb{E} \left\| \mathcal{C}_0^{-\frac{\rho}{2}} \mathcal{C}_1^{\frac{1}{2}} \mathcal{C}_1^{-\frac{1}{2}} \xi \right\|^2 \leq c \mathbb{E} \left\| \mathcal{C}_0^{\frac{\beta - \rho}{2}} \zeta \right\|^2 < \infty,$$

since $\beta - \rho > s_0$.

iv) By (ii) we have that $\mathcal{A}^{-1}u$ is μ_0-a.s. in the Cameron-Martin space of the Gaussian measures $\mathbb{P}$ and $\mathbb{P}_0$, thus the measures $\mathbb{P}$ and $\mathbb{P}_0$ are μ_0-a.s. equivalent [5, Theorem 2.8] and (iii) gives the result.

$\square$

2.2. *Guidelines For Applying The Theory.* The theory is naturally developed in the scale of Hilbert spaces defined via the prior. However application of the theory may be more natural in a different functional setting. We explain how the two may be connected. Let $\{\psi_k\}_{k\in\mathbb{N}}$ be an orthonormal basis of the separable Hilbert space $\mathcal{X}$. We define the spaces $\hat{X}^t$, $t \in \mathbb{R}$ as follows: for $t > 0$ we set

$$\hat{X}^t := \{u \in \mathcal{X} : \sum_{k=1}^{\infty} k^{2t}\langle u, \psi_k\rangle^2 < \infty\}$$

and the spaces $\hat{X}^{-t}$, $t > 0$ are defined by duality, $\hat{X}^{-t} := (\hat{X}^t)^*$.

For example, if we restrict ourselves to functions on a periodic domain $D = [0, L]^d$ and assume that $\{\psi_k\}_{k\in\mathbb{N}}$ is the Fourier basis of $\mathcal{X} = L^2(D)$, then the spaces $\hat{X}^t$ can be identified with the Sobolev spaces of periodic functions H^t, by rescaling: $H^t = \hat{X}^{\frac{t}{d}}$ [20, Proposition 5.39].

In the case $\sigma_0 < 1$, as explained in Remark 2.3 we have algebraic decay of the eigenvalues of $\mathcal{C}_0$ and in particular λ_k^2 decay like $k^{-\frac{1}{s_0}}$. If $\mathcal{C}_0$ is diagonalizable in the basis $\{\psi_k\}_{k\in\mathbb{N}}$, that is, if $\phi_k = \psi_k$, $k \in \mathbb{N}$, then it is straightforward to identify the spaces X^t with the spaces $\hat{X}^{\frac{t}{2s_0}}$. The advantage of this identification is that the spaces $\hat{X}^t$ do not depend on the prior so one can use them as a fixed reference point for expressing regularity, for example of the true solution.

In our subsequent analysis, we will require that the true solution lives in the Cameron-Martin space of the prior X^1, which in different choices of the prior (different σ_0) is a different space. Furthermore, we will assume that the true solution lives in X^γ for some $\gamma \geq 1$ and provide the convergence rate depending on the parameters γ, s_0, β, ℓ. The identification $X^\gamma = \hat{X}^{\frac{\gamma}{2s_0}}$ and the intuitive relation between the spaces $\hat{X}^t$ and the Sobolev spaces, enable us to understand the meaning of the assumptions on the true solution.

We can now formulate the following guidelines for applying the theory presented in the present paper: we work in a separable Hilbert space $\mathcal{X}$ with an orthonormal basis $\{\psi_k\}_{k\in\mathbb{N}}$ and we have some prior knowledge about the true solution $u^\dagger$ which can be expressed in terms of the spaces $\hat{X}^t$. The noise is assumed to be Gaussian $\mathcal{N}(0, \mathcal{C}_1)$, and the forward operator is known; that is, $\mathcal{C}_1$ and $\mathcal{A}^{-1}$ are known. We choose the prior $\mathcal{N}(0, \mathcal{C}_0)$, that is, we choose the covariance operator $\mathcal{C}_0$ and we can find the value of s_0. If the operator $\mathcal{C}_0$ is chosen to be diagonal in the basis $\{\psi_k\}_{k\in\mathbb{N}}$ then we can find the regularity of the true solution in terms of the spaces X^t, that is, the value of γ such that $u^\dagger \in X^\gamma$, and check that $\gamma \geq 1$ which is necessary for our theory to work. We then find the values of β and ℓ and calculate the value of Δ appearing in Assumption 2.4, checking that our choice of the prior is such that $\Delta > 2s_0$. We now have all the necessary information required for applying the Corollaries 6.4 and 6.5 presented in Section 6 to get the rate of convergence.

REMARK 2.6. *Observe that in the above mentioned example of periodic functions, we have the identification $X^1 = H^{\frac{d}{2s_0}}$, thus since $s_0 < 1$ we have that the assumption $u^\dagger \in X^1$ implies that $u^\dagger \in H^t$, for $t > \frac{d}{2}$. By the Sobolev embedding theorem [20, Theorem 5.31], this implies that the true solution is always assumed to be continuous. However, this is not a disadvantage of our method, since in many cases a Gaussian measure which charges $L^2(D)$ with probability one, can be shown to also charge the space of continuous functions with probability one [23, Lemma 6.25]*

3. Properties of the Posterior Mean and Covariance. We now make sense of the equation (1.7) weakly in the space X^1, under the assumptions presented in the previous section. To do so, we define the operator $\mathcal{B}_\lambda$ from (1.6) in X^1 and examine its properties. In Section 4 we demonstrate that (1.4) and (1.7) do indeed correspond to the posterior covariance and mean.

Consider the equation

(3.1) $$\mathcal{B}_\lambda w = r,$$

where

$$\mathcal{B}_\lambda = \mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1} + \lambda\mathcal{C}_0^{-1}.$$

Define the bilinear form $B \colon X^1 \times X^1 \to \mathbb{R}$,

$$B(u,v) := \left\langle \mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}u, \mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}v \right\rangle + \lambda\left\langle \mathcal{C}_0^{-\frac{1}{2}}u, \mathcal{C}_0^{-\frac{1}{2}}v \right\rangle, \ \ \forall u,v \in X^1.$$

DEFINITION 3.1. *Let $r \in X^{-1}$. An element $w \in X^1$ is called a weak solution of (3.1), if*

$$B(w,v) = \left\langle r, v \right\rangle, \ \forall v \in X^1.$$

PROPOSITION 3.2. *Under the Assumptions 2.4(1) and (2), for any $r \in X^{-1}$, there exists a unique weak solution $w \in X^1$ of (3.1).*

PROOF. We use the Lax-Milgram theorem in the Hilbert space X^1, since $r \in X^{-1} = (X^1)^*$.

i) $B \colon X^1 \times X^1 \to \mathbb{R}$ is coercive:

$$B(u,u) = \left\|\mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}u\right\|^2 + \lambda\left\|\mathcal{C}_0^{-\frac{1}{2}}u\right\|^2 \geq \lambda\|u\|_1^2, \ \forall u \in X^1.$$

ii) $B \colon X^1 \times X^1 \to \mathbb{R}$ is continuous: indeed by the Cauchy-Schwarz inequality and the Assumptions 2.4(1) and (2),

$$|B(u,v)| \leq \left\|\mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}u\right\|\left\|\mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}v\right\| + \lambda\left\|\mathcal{C}_0^{-\frac{1}{2}}u\right\|\left\|\mathcal{C}_0^{-\frac{1}{2}}v\right\|$$
$$\leq c\|u\|_{\beta-2\ell}\|v\|_{\beta-2\ell} + \lambda\|u\|_1\|v\|_1 \leq c'\|u\|_1\|v\|_1, \ \forall u,v \in X^1.$$

$\square$

REMARK 3.3. *The Lax-Milgram theorem defines a bounded operator $\mathcal{S} : X^{-1} \to X^1$, such that $B(\mathcal{S}r, v) = \langle r, v \rangle$ for all $v \in X^1$, which has a bounded inverse $\mathcal{S}^{-1} : X^1 \to X^{-1}$ such that $B(w, v) = \langle \mathcal{S}^{-1}w, v \rangle$ for all $v \in X^1$. Henceforward, we identify $\mathcal{B}_\lambda \equiv \mathcal{S}^{-1}$ and $\mathcal{B}_\lambda^{-1} \equiv \mathcal{S}$. Furthermore, note that in Proposition 3.2, Lemma 3.4 below, and the three propositions in Section 5, we only require $\Delta > 0$ and not the stronger assumption $\Delta > 2s_0$. However, in all our other results we actually need $\Delta > 2s_0$.*

LEMMA 3.4. *Suppose the Assumptions 2.4(1) and (2) hold. Then the operator $\mathcal{S}^{-1} = \mathcal{B}_\lambda : X^1 \to X^{-1}$ is identical to the operator $\mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1} + \lambda\mathcal{C}_0^{-1} : X^1 \to X^{-1}$, where $\mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1}$ is defined weakly in $X^{\beta-2\ell}$.*

PROOF. The Lax-Milgram theorem implies that $\mathcal{B}_\lambda : X^1 \to X^{-1}$ is bounded. Moreover, $\mathcal{C}_0^{-1} : X^1 \to X^{-1}$ is bounded, thus the operator $K := \mathcal{B}_\lambda - \lambda\mathcal{C}_0^{-1} : X^1 \to X^{-1}$ is also bounded and satisfies

$$(3.2) \qquad \langle Ku, v \rangle = \langle \mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}u, \mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}v \rangle, \ \forall u, v \in X^1.$$

Define $\mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1}$ weakly in $X^{\beta-2\ell}$, by the bilinear form $A : X^{\beta-2\ell} \times X^{\beta-2\ell} \to \mathbb{R}$ given by

$$A(u, v) = \langle \mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}u, \mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}v \rangle, \ \forall u, v \in X^{\beta-2\ell}.$$

By Assumption 2.4(2), A is coercive and continuous in $X^{\beta-2\ell}$, thus by the Lax-Milgram theorem, there exists a uniquely defined, boundedly invertible, operator $T : X^{2\ell-\beta} \to X^{\beta-2\ell}$ such that $A(u, v) = \langle T^{-1}u, v \rangle$ for all $v \in X^{\beta-2\ell}$. We identify $\mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1}$ with the bounded operator $T^{-1} : X^{\beta-2\ell} \to X^{2\ell-\beta}$. By Assumption 2.4(1) we have $\Delta > 0$ hence

$$\left\|\mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1}u\right\|_{-1} \leq c\left\|\mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1}u\right\|_{2\ell-\beta} \leq c\|u\|_{\beta-2\ell} \leq c\|u\|_1, \ \forall u \in X^1,$$

that is, $\mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1} : X^1 \to X^{-1}$ is bounded. By the definition of $T^{-1} = \mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1}$ and (3.2), this implies that $K = \mathcal{B}_\lambda - \lambda\mathcal{C}_0^{-1} = \mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1}$. $\square$

PROPOSITION 3.5. *Under the Assumptions 2.1, 2.4(1),(2),(3),(6), there exists a unique weak solution, $m \in X^1$ of equation (1.7), $\nu(du, dy)$-almost surely.*

PROOF. It suffices to show that $\mathcal{A}^{-1}\mathcal{C}_1^{-1}y \in X^{-1}$, $\nu(du, dy)$-almost surely. Indeed, by Lemma 2.5(iv) we have that $y \in X^\rho$ $\nu(du, dy)$-a.s. for all $\rho < \beta - s_0$, thus by the Assumption 2.4(6)

$$\left\|\mathcal{C}_0^{\frac{1}{2}}\mathcal{A}^{-1}\mathcal{C}_1^{-1}y\right\| \leq c\left\|\mathcal{C}_0^{\frac{1}{2}+\ell-\beta}y\right\| < \infty,$$

since $2\beta - 2\ell - 1 < \beta - s_0$, which holds by the Assumption 2.4(1). $\square$

4. Posterior Identification. Suppose that in the problem (1.1) we have $u \sim \mu_0 = \mathcal{N}(0, \mathcal{C}_0)$ and $\xi \sim \mathcal{N}(0, \mathcal{C}_1)$, where u is independent of ξ. Then we have that $y|u \sim \mathbb{P} = \mathcal{N}(\mathcal{A}^{-1}u, \frac{1}{n}\mathcal{C}_1)$. Let μ^y be the posterior measure on $u|y$.

In this section we prove a number of facts concerning the posterior measure μ^y for $u|y$. First, in Theorem 4.1 we prove that this measure has density with respect to the prior measure μ_0, identify this density and show that μ^y is Lipschitz in y, with respect to the Hellinger metric. Continuity in y will require the introduction of the space $X^{s+\beta-2\ell}$, to which u drawn from μ_0 belongs almost surely. Secondly, in Theorem 4.2, we show that μ^y is Gaussian and identify the covariance and mean via equations (1.4) and (1.7). This identification will form the basis for our analysis of posterior contraction in the following section.

THEOREM 4.1. *Under the Assumptions 2.1, 2.4(1),(2),(3),(4),(5), the posterior measure μ^y is absolutely continuous with respect to μ_0 and*

$$(4.1) \qquad \frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z(y)} \exp(-\Phi(u, y)),$$

where

$$(4.2) \qquad \Phi(u, y) := \frac{n}{2} \left\| \mathcal{C}_1^{-\frac{1}{2}} \mathcal{A}^{-1}u \right\|^2 - n \langle \mathcal{C}_1^{-\frac{1}{2}} y, \mathcal{C}_1^{-\frac{1}{2}} \mathcal{A}^{-1}u \rangle$$

and $Z(y) \in (0, \infty)$ is the normalizing constant. Furthermore, the map $y \mapsto \mu^y$ is Lipschitz continuous, with respect to the Hellinger metric: let $s = s_0 + \varepsilon$, $0 < \varepsilon < (\Delta - 2s_0) \wedge \sigma_0$; then there exists $c = c(r)$ such that for all $y, y' \in X^{\beta-s}$ with $\|y\|_{\beta-s}, \|y'\|_{\beta-s} \leq r$

$$d_{\mathrm{Hell}}(\mu^y, \mu^{y'}) \leq c \|y - y'\|_{\beta-s}.$$

Consequently, the μ^y-expectation of any polynomially bounded function $f: X^{s+\beta-2\ell} \to E$, where $(E, \|\cdot\|_E)$ is a Banach space, is locally Lipschitz continuous in y. In particular, the posterior mean is locally Lipschitz continuous in y as a function $X^{\beta-s} \to X^{s+\beta-2\ell}$.

THEOREM 4.2. *Under the Assumptions 2.1, 2.4, the posterior measure $\mu^y(du)$ is Gaussian $\mu^y = \mathcal{N}(m, \mathcal{C})$, where $\mathcal{C}$ is given by (1.4) and m is a weak solution of (1.7).*

The proofs of these two theorems are presented in the next two sections. Each proof is based on a series of lemmas.

4.1. *Proof of Theorem 4.1.* In this subsection we prove Theorem 4.1. We first prove several useful estimates regarding Φ defined in (4.2), for $u \in X^{s+\beta-2\ell}$ and $y \in X^{\beta-s}$, where $s \in (s_0, 1]$. Observe that, under the Assumptions 2.1, 2.4(1),(2),(3), for $s = s_0 + \varepsilon$ where $\varepsilon > 0$ sufficiently small, the Lemma 2.5 implies on the one hand that $u \in X^{s+\beta-2\ell}$ $\mu_0(du)$-almost surely and on the other hand that $y \in X^{\beta-s}$ $\nu(du, dy)$-almost surely.

LEMMA 4.3.   *Under the Assumptions 2.1, 2.4(2),(4),(5), for any $s \in (s_0, 1]$, the potential Φ given by (4.2) satisfies:*

i) *for every $\delta > 0$ and $r > 0$, there exists an $M = M(\delta, r) \in \mathbb{R}$, such that for all $u \in X^{s+\beta-2\ell}$ and all $y \in X^{\beta-s}$ with $\|y\|_{\beta-s} \leq r$,*

$$\Phi(u, y) \geq M - \delta \|u\|^2_{s+\beta-2\ell};$$

ii) *for every $r > 0$, there exists a $K = K(r) > 0$, such that for all $u \in X^{s+\beta-2\ell}$ and $y \in X^{\beta-s}$ with $\|u\|_{s+\beta-2\ell}, \|y\|_{\beta-s} \leq r$,*

$$\Phi(u, y) \leq K;$$

iii) *for every $r > 0$, there exists an $L = L(r) > 0$, such that for all $u_1, u_2 \in X^{s+\beta-2\ell}$ and $y \in X^{\beta-s}$ with $\|u_1\|_{s+\beta-2\ell}, \|u_2\|_{s+\beta-2\ell}, \|y\|_{\beta-s} \leq r$,*

$$|\Phi(u_1, y) - \Phi(u_2, y)| \leq L \|u_1 - u_2\|_{s+\beta-2\ell};$$

iv) *for every $\delta > 0$ and $r > 0$, there exists an $c = c(\delta, r) \in \mathbb{R}$, such that for all $y_1, y_2 \in X^{\beta-s}$ with $\|y_1\|_{\beta-s}, \|y_2\|_{\beta-s} \leq r$ and for all $u \in X^{s+\beta-2\ell}$,*

$$|\Phi(u, y_1) - \Phi(u, y_2)| \leq \exp\left(\delta \|u\|^2_{s+\beta-2\ell} + c\right) \|y_1 - y_2\|_{\beta-s}.$$

PROOF.

i) By first using the Cauchy-Schwarz inequality, then the Assumptions 2.4 (4) and (5), and then the Cauchy with δ' inequality for $\delta' > 0$ sufficiently small, we have

$$\Phi(u, y) = \frac{n}{2} \left\| \mathcal{C}_1^{-\frac{1}{2}} \mathcal{A}^{-1} u \right\|^2 - n \langle \mathcal{C}_0^{\frac{s}{2}} \mathcal{C}_1^{-\frac{1}{2}} y, \mathcal{C}_0^{-\frac{s}{2}} \mathcal{C}_1^{-\frac{1}{2}} \mathcal{A}^{-1} u \rangle$$

$$\geq -n \|\mathcal{C}_0^{\frac{s}{2}} \mathcal{C}_1^{-\frac{1}{2}} y\| \|\mathcal{C}_0^{-\frac{s}{2}} \mathcal{C}_1^{-\frac{1}{2}} \mathcal{A}^{-1} u\| \geq -cn \|y\|_{\beta-s} \|u\|_{s+\beta-2\ell}$$

$$\geq -\frac{cn}{4\delta'} \|y\|^2_{\beta-s} - cn\delta' \|u\|^2_{s+\beta-2\ell} \geq M(r, \delta) - \delta \|u\|^2_{s+\beta-2\ell}.$$

ii) By the Cauchy-Schwarz inequality and the Assumptions 2.4(2),(4) and (5), we have since $s > s_0 \geq 0$

$$\Phi(u, y) \leq \frac{n}{2} \left\| \mathcal{C}_1^{-\frac{1}{2}} \mathcal{A}^{-1} u \right\|^2 + n \|\mathcal{C}_0^{\frac{s}{2}} \mathcal{C}_1^{-\frac{1}{2}} y\| \|\mathcal{C}_0^{-\frac{s}{2}} \mathcal{C}_1^{-\frac{1}{2}} \mathcal{A}^{-1} u\|$$

$$\leq c \frac{n}{2} \|u\|^2_{\beta-2\ell} + cn \|y\|_{\beta-s} \|u\|_{s+\beta-2\ell} \leq K(r).$$

iii) By first using the Assumptions 2.4 (4) and (5) and the triangle inequality, and then the Assumption 2.4(2) and the reverse triangle inequality, we have since $s > s_0 \geq 0$

$$|\Phi(u_1, y) - \Phi(u_2, y)| =$$

$$\frac{n}{2} \left| \left\| \mathcal{C}_1^{-\frac{1}{2}} \mathcal{A}^{-1} u_1 \right\|^2 - \left\| \mathcal{C}_1^{-\frac{1}{2}} \mathcal{A}^{-1} u_2 \right\|^2 + 2 \langle \mathcal{C}_0^{\frac{s}{2}} \mathcal{C}_1^{-\frac{1}{2}} y, \mathcal{C}_0^{-\frac{s}{2}} \mathcal{C}_1^{-\frac{1}{2}} \mathcal{A}^{-1} (u_2 - u_1) \rangle \right|$$

$$\leq \frac{n}{2}\left|\|\mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}u_1\|^2 - \|\mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}u_2\|^2\right| + cn\|y\|_{\beta-s}\|u_1 - u_2\|_{s+\beta-2\ell}$$

$$\leq cn\|u_1 - u_2\|_{\beta-2\ell}\left(\|u_1\|_{\beta-2\ell} + \|u_2\|_{\beta-2\ell}\right) + cnr\|u_1 - u_2\|_{s+\beta-2\ell}$$

$$\leq L(r)\|u_1 - u_2\|_{s+\beta-2\ell}.$$

iv) By first using the Cauchy-Schwarz inequality and then the Assumptions 2.4(4) and (5), we have

$$|\Phi(u, y_1) - \Phi(u, y_2)| = n\left|\langle\mathcal{C}_0^{\frac{s}{2}}\mathcal{C}_1^{-\frac{1}{2}}(y_1 - y_2), \mathcal{C}_0^{-\frac{s}{2}}\mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}u\rangle\right|$$

$$\leq n\left\|\mathcal{C}_0^{\frac{s}{2}}\mathcal{C}_1^{-\frac{1}{2}}(y_1 - y_2)\right\|\left\|\mathcal{C}_0^{-\frac{s}{2}}\mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}u\right\|$$

$$\leq cn\|y_1 - y_2\|_{\beta-s}\|u\|_{s+\beta-2\ell}$$

$$\leq \exp\left(\delta\|u\|_{s+\beta-2\ell}^2 + c\right)\|y_1 - y_2\|_{\beta-s}.$$

$\square$

COROLLARY 4.4.   *Under the Assumptions 2.1, 2.4(1),(2),(4),(5)*

$$Z(y) := \int_{\mathcal{X}} \exp(-\Phi(u, y))\mu_0(du) > 0,$$

for all $y \in X^{\beta-s}, s = s_0 + \varepsilon$ where $0 < \varepsilon < (\Delta - 2s_0) \wedge \sigma_0$. In particular, if in addition the Assumption 2.4(3) holds, then $Z(y) > 0$ ν-almost surely.

PROOF. Fix $y \in X^{\beta-s}$ and set $r = \|y\|_{\beta-s}$. Gaussian measures on separable Hilbert spaces are full [5, Proposition 1.25], hence since by Lemma 2.5(i) $\mu_0(X^{s+\beta-2\ell}) = 1$, we have that $\mu_0(B_{X^{s+\beta-2\ell}}(r)) > 0$. By Lemma 4.3(ii), there exists $K(r) > 0$ such that

$$\int_{\mathcal{X}} \exp(-\Phi(u, y))\mu_0(du) \geq \int_{B_{X^{s+\beta-2\ell}}(r)} \exp(-\Phi(u, y))\mu_0(du)$$

$$\geq \int_{B_{X^{s+\beta-2\ell}}(r)} \exp(-K(r))\mu_0(du) > 0.$$

Recalling that, under the additional Assumption 2.4(3), by Lemma 2.5(iv) we have $y \in X^{\beta-s}$ ν-almost surely for all $s > s_0$, completes the proof. $\square$

We are now ready to prove Theorem 4.1:

PROOF OF THEOREM 4.1. Recall that $\nu_0 = \mathbb{P}_0(dy)\otimes\mu_0(du)$ and $\nu = \mathbb{P}(dy|u)\mu_0(du)$. By the Cameron-Martin formula [3, Corollary 2.4.3], since by Lemma 2.5(ii) we have $\mathcal{A}^{-1}u \in \mathcal{D}(\mathcal{C}_1^{-\frac{1}{2}})$ μ_0-a.s., we get for μ_0-almost all u

$$\frac{d\mathbb{P}}{d\mathbb{P}_0}(y|u) = \exp(-\Phi(u, y)),$$

thus we have for μ_0-almost all u

$$\frac{d\nu}{d\nu_0}(y, u) = \exp(-\Phi(u, y)).$$

By [11, Lemma 5.3] and Corollary 4.4 we have the relation (4.1).

For the proof of the Lipschitz continuity of the posterior measure in y, with respect to the Hellinger distance, we apply [23, Theorem 4.2] for $Y = X^{\beta-s}, X = X^{s+\beta-2\ell}$, using Lemma 4.3 and the fact that $\mu_0(X^{s+\beta-2\ell}) = 1$, by Lemma 2.5(i).   □

4.2. *Proof of Theorem 4.2.*   We first give an overview of the proof of Theorem 4.2. Let $y|u \sim \mathbb{P} = \mathcal{N}(\mathcal{A}^{-1}u, \frac{1}{n}\mathcal{C}_1)$ and $u \sim \mu_0$. Then by Proposition 3.5, there exists a unique weak solution, $m \in X^1$, of (1.7), $\nu(du, dy)$-almost surely. That is, with $\nu(du, dy)$-probability equal to one, there exists an $m = m(y) \in X^1$ such that

$$B(m, v) = b^y(v), \ \ \forall v \in X^1,$$

where the bilinear form B is defined in Section 3, and $b^y(v) = \langle \mathcal{A}^{-1}\mathcal{C}_1^{-1}y, v \rangle$. In the following we show that $\mu^y = \mathcal{N}(m, \mathcal{C})$, where

$$\mathcal{C}^{-1} = n\mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1} + \frac{1}{\tau^2}\mathcal{C}_0^{-1}.$$

The proof has the same structure as the proof for the identification of the posterior in [19]. We define the Gaussian measure $\mathcal{N}(m^N, \mathcal{C}^N)$, which is the independent product of a measure identical to $\mathcal{N}(m, \mathcal{C})$ in the finite-dimensional space $\mathcal{X}^N$ spanned by the first N eigenfunctions of $\mathcal{C}_0$, and a measure identical to μ_0 in $(\mathcal{X}^N)^\perp$. We next show that $\mathcal{N}(m^N, \mathcal{C}^N)$ converges weakly to the measure μ^y which as a weak limit of Gaussian measures has to be Gaussian $\mu^y = \mathcal{N}(\overline{m}, \overline{\mathcal{C}})$, and we then identify $\overline{m}$ and $\overline{\mathcal{C}}$ with m, $\mathcal{C}$ respectively.

Fix y drawn from ν and let P^N be the orthogonal projection of $\mathcal{X}$ to the finite-dimensional space $\text{span}\{\phi_1, ..., \phi_N\} := \mathcal{X}^N$, where as in Section 2, $\{\phi_k\}_{k=1}^\infty$ is an orthonormal eigenbasis of $\mathcal{C}_0$ in $\mathcal{X}$. Let $Q^N = I - P^N$. We define $\mu^{N,y}$ by

(4.3) $$\frac{d\mu^{N,y}}{d\mu_0}(u) = \frac{1}{Z^N(y)} \exp(-\Phi^N(u, y))$$

where $\Phi^N(u, y) := \Phi(P^N u, y)$ and

$$Z^N(y) := \int_\mathcal{X} \exp(-\Phi^N(u, y))\mu_0(du).$$

LEMMA 4.5.   *We have* $\mu^{N,y} = \mathcal{N}(m^N, \mathcal{C}^N)$, *where*

$$P^N\mathcal{C}^{-1}P^N m^N = nP^N\mathcal{A}^{-1}\mathcal{C}_1^{-1}y,$$

$$P^N\mathcal{C}^N P^N = P^N\mathcal{C}P^N, \ Q^N\mathcal{C}^N Q^N = \tau^2 Q^N\mathcal{C}_0 Q^N$$

and $P^N\mathcal{C}^N Q^N = Q^N\mathcal{C}^N P^N = 0$.

PROOF. Let $u \in \mathcal{X}^N$. Since $u = P^N u$ we have by (4.3)

$$d\mu^{N,y}(P^N u) \propto \exp\left(-\Phi(P^N u; y)\right) d\mu_0(P^N u).$$

The right hand side is N-dimensional Gaussian with density proportional to the exponential of the following expression

$$(4.4) \qquad -\frac{n}{2}\left\|\mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}P^N u\right\|^2 + n\left\langle\mathcal{C}_1^{-\frac{1}{2}}y, \mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}P^N u\right\rangle - \frac{1}{2\tau^2}\left\|\mathcal{C}_0^{-\frac{1}{2}}P^N u\right\|^2,$$

which by completing the square we can write as

$$-\frac{1}{2}\left\|(\tilde{\mathcal{C}}^N)^{-\frac{1}{2}}(u - \tilde{m}^N)\right\|^2 + c(y),$$

where $\tilde{\mathcal{C}}^N$ is the covariance matrix and $\tilde{m}^N$ the mean. By equating with expression (4.4), we find that $(\tilde{\mathcal{C}}^N)^{-1} = P^N \mathcal{C}^{-1} P^N$ and $(\tilde{\mathcal{C}}^N)^{-1}\tilde{m}^N = nP^N\mathcal{A}^{-1}\mathcal{C}_1^{-1}y$, thus on $\mathcal{X}^N$ we have that $\mu^{N,y} = \mathcal{N}(\tilde{m}^N, \tilde{\mathcal{C}}^N)$. On $(\mathcal{X}^N)^\perp$, the Radon-Nikodym derivative in (4.3) is equal to 1, hence $\mu^{N,y} = \mu_0 = \mathcal{N}(0, \tau^2 \mathcal{C}_0)$. $\square$

PROPOSITION 4.6. *Under the Assumptions 2.1, 2.4(1),(2),(3),(4),(5), for all $y \in X^{\beta-s}$, $s = s_0 + \varepsilon$, where $0 < \varepsilon < (\Delta - 2s_0) \wedge \sigma_0$, the measures $\mu^{N,y}$ converge weakly in $\mathcal{X}$ to μ^y, where μ^y is defined in Theorem 4.1. In particular, $\mu^{N,y}$ converge weakly in $\mathcal{X}$ to μ^y ν-almost surely.*

PROOF. Fix $y \in X^{\beta-s}$. Let $f : \mathcal{X} \to \mathbb{R}$ be continuous and bounded. Then by (4.1), (4.3) and Lemma 2.5(i), we have that

$$\int_{\mathcal{X}} f(u)\mu^{N,y}(du) = \frac{1}{Z^N}\int_{X^{s+\beta-2\ell}} f(u)e^{-\Phi^N(u,y)}\mu_0(du)$$

and

$$\int_{\mathcal{X}} f(u)\mu^y(du) = \frac{1}{Z}\int_{X^{s+\beta-2\ell}} f(u)e^{-\Phi(u,y)}\mu_0(du).$$

Let $u \in X^{s+\beta-2\ell}$ and set $r_1 = \max\{\|u\|_{s+\beta-2\ell}, \|y\|_{\beta-s}\}$ to get, by Lemma 4.3(iii), that $\Phi^N(u,y) \to \Phi(u,y)$, since $\left\|P^N u\right\|_{s+\beta-2\ell} \leq \|u\|_{s+\beta-2\ell} \leq r_1$. By Lemma 4.3(i), for any $\delta > 0$, for $r_2 = \|y\|_{\beta-s}$, there exists $M(\delta, r_2) \in \mathbb{R}$ such that

$$\left|f(u)e^{-\Phi^N(u,y)}\right| \leq \|f\|_\infty e^{\delta\|u\|^2_{s+\beta-2\ell} - M(\delta, r_2)}, \ \forall u \in X^{s+\beta-2\ell},$$

where the right hand side is μ_0-integrable for δ sufficiently small by the Fernique Theorem [3, Theorem 2.8.5]. Hence, by the Dominated Convergence Theorem, we have that $\int_{\mathcal{X}} f(u)\mu^{N,y}(du) \to \int_{\mathcal{X}} f(u)\mu^y(du)$, as $N \to \infty$, where we get the convergence of the constants $Z^N \to Z$ by choosing $f \equiv 1$. Thus we have $\mu^{N,y} \Rightarrow \mu^y$. Recalling, that $y \in X^{\beta-s}$ ν-almost surely completes the proof. $\square$

We are now ready to prove Theorem 4.2:

PROOF OF THEOREM 4.2. By Proposition 4.6 we have that $\mu^{N,y}$ converge weakly in $\mathcal{X}$ to the measure μ^y, ν-almost surely. Since by Lemma 4.5, the measures $\mu^{N,y}$ are Gaussian, the limiting measure μ^y is also Gaussian. To see this we argue as follows. The weak convergence of measures implies the pointwise convergence of the Fourier transforms of the measures, thus by Levy's continuity theorem [13, Theorem 4.3] all the one dimensional projections of $\mu^{N,y}$, which are Gaussian, converge weakly to the corresponding one dimensional projections of μ^y. By the fact that the class of Gaussian distributions in $\mathbb{R}$ is closed under weak convergence [13, Chapter 4, Exercise 2], we get that all the one dimensional projections of the μ^y are Gaussian, thus μ^y is a Gaussian measure in $\mathcal{X}$, $\mu^y = \mathcal{N}(\overline{m}, \overline{\mathcal{C}})$ for some $\overline{m} \in \mathcal{X}$ and a self-adjoint, positive semi definite, trace class linear operator $\overline{\mathcal{C}}$. It suffices to show that $\overline{m} = m$ and $\overline{\mathcal{C}} = \mathcal{C}$.

We use the standard Galerkin method to show that $m^N \to m$ in $\mathcal{X}$. Indeed, since by their definition m^N solve (1.7) in the N-dimensional spaces $\mathcal{X}^N$, for $e = m - m^N$, we have that $B(e, v) = 0$, $\forall v \in \mathcal{X}^N$. By the coercivity and the continuity of B (see Proposition 3.2)

$$\left\|e\right\|_1^2 \leq cB(e, e) = cB(e, m - z) \leq c\left\|e\right\|_1 \left\|m - z\right\|_1, \ \forall z \in \mathcal{X}^N.$$

Choose $z = P^N m$ to obtain

$$\left\|m - m^N\right\| \leq c\left\|m - P^N m\right\|_1,$$

where as $N \to \infty$ the right hand side converges to zero since $m \in X^1$. On the other hand, by [3, Example 3.8.15], we have that $m^N \to \overline{m}$ in $\mathcal{X}$, hence we conclude that $\overline{m} = m$, as required.

For the identification of the covariance operator, note that by the definition of $\mathcal{C}^N$ we have

$$\mathcal{C}^N = P^N \mathcal{C} P^N + (I - P^N)\mathcal{C}_0(I - P^N).$$

Recall that $\{\phi_k\}_{k=1}^\infty$ are the eigenfunctions of $\mathcal{C}_0$ and fix $k \in \mathbb{N}$. Then, for $N > k$ and any $w \in \mathcal{X}$, we have that

$$\left|\langle w, \mathcal{C}^N \phi_k \rangle - \langle w, \mathcal{C}\phi_k \rangle\right| = \left|\langle w, (P^N - I)\mathcal{C}\phi_k \rangle\right|$$
$$\leq \left\|(P^N - I)w\right\|\left\|\mathcal{C}\phi_k\right\|,$$

where the right hand side converges to zero as $N \to \infty$, since $w \in \mathcal{X}$. This implies that $\mathcal{C}^N \phi_k$ converges to $\mathcal{C}\phi_k$ weakly in $\mathcal{X}$, as $N \to \infty$ and this holds for any $k \in \mathbb{N}$. On the other hand by [3, Example 3.8.15], we have that $\mathcal{C}^N \phi_k \to \overline{\mathcal{C}}\phi_k$ in $\mathcal{X}$, as $N \to \infty$, for all $k \in \mathbb{N}$. It follows that $\overline{\mathcal{C}}\phi_k = \mathcal{C}\phi_k$, for every k and since $\{\phi_k\}_{k=1}^\infty$ is an orthonormal basis of $\mathcal{X}$, we have that $\overline{\mathcal{C}} = \mathcal{C}$. $\qquad\square$

5. Operator norm bounds on $\mathcal{B}_\lambda^{-1}$. The following propositions contain several operator norm estimates on the inverse of $\mathcal{B}_\lambda$ and related quantities, and in particular estimates on the singular dependence of this operator as $\lambda \to 0$. These are the key tools used in Section 6 to obtain posterior contraction results. In all of them we make use of the interpolation inequality in Hilbert scales, [7, Proposition 8.19]. Recall that we consider $\mathcal{B}_\lambda$ defined on X^1, as explained in Remark 3.3.

PROPOSITION 5.1.   *Let $\eta_1 = (1 - \theta_1)(\beta - 2\ell) + \theta_1$, where $\theta_1 \in [0, 1]$. Under the Assumptions 2.4(1),(2) and (6) the following operator norm bounds hold: there is $c > 0$ independent of θ_1 such that*

$$\left\|\mathcal{B}_\lambda^{-1}\mathcal{A}^{-1}\mathcal{C}_1^{-1}\right\|_{\mathcal{L}(X^{2\beta-2\ell-\eta_1}, X^{\beta-2\ell})} \leq c\lambda^{-\frac{\theta_1}{2}}$$

and

$$\left\|\mathcal{B}_\lambda^{-1}\mathcal{A}^{-1}\mathcal{C}_1^{-1}\right\|_{\mathcal{L}(X^{2\beta-2\ell-\eta_1}, X^1)} \leq c\lambda^{-\frac{\theta_1+1}{2}}.$$

PROOF. Let $h \in X^{2\beta-2\ell-\eta_1} = X^{\beta-\theta_1\Delta}$. Then $\mathcal{A}^{-1}\mathcal{C}_1^{-1}h \in X^{-1}$. Indeed, by Assumption 2.4(6) for $\eta = 1$,

$$\left\|\mathcal{C}_0^{\frac{1}{2}}\mathcal{A}^{-1}\mathcal{C}_1^{-1}h\right\| \leq c\left\|\mathcal{C}_0^{\frac{1}{2}+\ell-\beta}h\right\| = c\left\|\mathcal{C}_0^{\frac{\Delta-\beta}{2}}h\right\| = c\|h\|_{\beta-\Delta} \leq c\|h\|_{\beta-\theta_1\Delta}.$$

By Proposition 3.2 for $r = \mathcal{A}^{-1}\mathcal{C}_1^{-1}h$, there exists a unique weak solution of (3.1), $z \in X^1$. By Definition 3.1, for $v = z \in X^1$, we get

$$\left\|\mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}z\right\|^2 + \lambda\left\|\mathcal{C}_0^{-\frac{1}{2}}z\right\|^2 = \langle\mathcal{C}_0^{\frac{\eta_1}{2}}\mathcal{A}^{-1}\mathcal{C}_1^{-1}h, \mathcal{C}_0^{-\frac{\eta_1}{2}}z\rangle.$$

Using the Assumptions 2.4(2) and (6), and the Cauchy-Schwarz inequality, we get

$$\|z\|_{\beta-2\ell}^2 + \lambda\|z\|_1^2 \leq c\left\|\mathcal{C}_0^{\frac{\eta_1}{2}+\ell-\beta}h\right\|\|z\|_{\eta_1}.$$

We interpolate the norm on z appearing on the right hand side between the norms on z appearing on the left hand side, then use the Cauchy with ε inequality, and then Young's inequality for $p = \frac{1}{1-\theta_1}, q = \frac{1}{\theta_1}$, to get successively, for $c > 0$ a changing constant

$$\|z\|_{\beta-2\ell}^2 + \lambda\|z\|_1^2 \leq c\left\|\mathcal{C}_0^{\frac{\eta_1}{2}+\ell-\beta}h\right\|\|z\|_{\beta-2\ell}^{1-\theta_1}\lambda^{-\frac{\theta_1}{2}}\left(\lambda^{\frac{1}{2}}\|z\|_1\right)^{\theta_1}$$

$$\leq \frac{c}{2\varepsilon}\left(\lambda^{-\theta_1}\left\|\mathcal{C}_0^{\frac{\eta_1}{2}+\ell-\beta}h\right\|^2\right) + \frac{c\varepsilon}{2}\left(\|z\|_{\beta-2\ell}^{2(1-\theta_1)}\left(\lambda^{\frac{1}{2}}\|z\|_1\right)^{2\theta_1}\right)$$

$$\leq \frac{c}{2\varepsilon}\left(\lambda^{-\theta_1}\left\|\mathcal{C}_0^{\frac{\eta_1}{2}+\ell-\beta}h\right\|^2\right) + \frac{c\varepsilon}{2}\left((1-\theta_1)\|z\|_{\beta-2\ell}^2 + \theta_1\lambda\|z\|_1^2\right).$$

By choosing $\varepsilon > 0$ small enough we get, for $c > 0$ independent of θ, λ,

$$\|z\|_{\beta-2\ell} \leq c\lambda^{-\frac{\theta_1}{2}}\left\|\mathcal{C}_0^{\frac{\eta_1}{2}+\ell-\beta}h\right\| \quad \text{and} \quad \|z\|_1 \leq c\lambda^{-\frac{\theta_1+1}{2}}\left\|\mathcal{C}_0^{\frac{\eta_1}{2}+\ell-\beta}h\right\|.$$

Replacing $z = \mathcal{B}_\lambda^{-1}\mathcal{A}^{-1}\mathcal{C}_1^{-1}h$ gives the result.   □

PROPOSITION 5.2.   *Let $\eta_2 = (1 - \theta_2)(\beta - 2\ell) + \theta_2$, where $\theta_2 \in [0, 1]$. Under the Assumptions 2.4(1) and (2), the following operator norm bounds hold: there is $c > 0$ independent of θ_2 such that*

$$\left\|\mathcal{B}_\lambda^{-1}\mathcal{C}_0^{-1}\right\|_{\mathcal{L}(X^{2-\eta_2}, X^{\beta-2\ell})} \leq c\lambda^{-\frac{\theta_2}{2}}$$

and

$$\left\|\mathcal{B}_\lambda^{-1}\mathcal{C}_0^{-1}\right\|_{\mathcal{L}(X^{2-\eta_2}, X^1)} \leq c\lambda^{-\frac{\theta_2+1}{2}}.$$

PROOF. Let $h \in X^{2-\eta_2}$. Then $\mathcal{C}_0^{-1}h \in X^{-1}$, since $\eta_2 \leq 1$. By Proposition 3.2 for $r = \mathcal{C}_0^{-1}h$, there exists a unique weak solution of (3.1), $z \in X^1$. By Definition 3.1, for $v = z \in X^1$, we have

$$\left\|\mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}z\right\|^2 + \lambda\left\|\mathcal{C}_0^{-\frac{1}{2}}z\right\|^2 = \left\langle \mathcal{C}_0^{\frac{\eta_2}{2}-1}h, \mathcal{C}_0^{-\frac{\eta_2}{2}}z\right\rangle,$$

and by the Assumption 2.4(2) and the Cauchy-Schwarz inequality, we get

$$\left\|z\right\|_{\beta-2\ell}^2 + \lambda\left\|z\right\|_1^2 \leq c\left\|\mathcal{C}_0^{\frac{\eta_2}{2}-1}h\right\|\left\|z\right\|_{\eta_2}.$$

We interpolate the norm on z appearing on the right hand side between the norms on z appearing on the left hand side to get as in the proof of Proposition 5.1, for $c > 0$ independent of θ, λ,

$$\left\|z\right\|_{\beta-2\ell} \leq c\lambda^{-\frac{\theta_2}{2}}\left\|\mathcal{C}_0^{\frac{\eta_2}{2}-1}h\right\| \quad \text{and} \quad \left\|z\right\|_1 \leq c\lambda^{-\frac{\theta_2+1}{2}}\left\|\mathcal{C}_0^{\frac{\eta_2}{2}-1}h\right\|.$$

Replacing $z = \mathcal{B}_\lambda^{-1}\mathcal{C}_0^{-1}h$ gives the result. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

PROPOSITION 5.3. Let $\eta_3 = (1-\theta_3)(\beta - 2\ell - s) + \theta_3(1-s)$, where $\theta_3 \in [0,1]$ and $s \in (s_0, 1]$, where $s_0 \in [0,1)$ as defined in Lemma 2.2. Under the Assumptions 2.4(1) and (2), the following norm bounds hold: there is $c > 0$ independent of θ_3 such that

$$\left\|\mathcal{C}_0^{-\frac{s}{2}}\mathcal{B}_\lambda^{-1}\mathcal{C}_0^{-\frac{s}{2}}\right\|_{\mathcal{L}(X^{\beta-2\ell-s}, X^{-\eta_3})} \leq c\lambda^{-\frac{\theta_3}{2}}$$

and

$$\left\|\mathcal{C}_0^{-\frac{s}{2}}\mathcal{B}_\lambda^{-1}\mathcal{C}_0^{-\frac{s}{2}}\right\|_{\mathcal{L}(X^{1-s}, X^{-\eta_3})} \leq c\lambda^{-\frac{\theta_3+1}{2}}.$$

Furthermore,

$$\left\|\mathcal{C}_0^{-\frac{s}{2}}\mathcal{B}_\lambda^{-1}\mathcal{C}_0^{-\frac{s}{2}}\right\|_{\mathcal{L}(X)} \leq c\lambda^{-\frac{2\ell-\beta+s}{\Delta}}, \quad \forall s \in (\{\beta - 2\ell\} \vee s_0, 1].$$

PROOF. Let $h \in X^{-\eta_3} = X^{(1-\theta_3)\Delta+s-1}$. Then $h \in X^{s-1}$, since $\Delta > 0$, thus $\mathcal{C}_0^{-\frac{s}{2}}h \in X^{-1}$. By Proposition 3.2 for $r = \mathcal{C}_0^{-\frac{s}{2}}h$, there exists a unique weak solution of (3.1), $z' \in X^1$. Since for $v \in X^{1-s}$ we have that $\mathcal{C}_0^{\frac{s}{2}}v \in X^1$, we conclude that for any $v \in X^{1-s}$

$$\left\langle \mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}\mathcal{C}_0^{\frac{s}{2}}z, \mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}\mathcal{C}_0^{\frac{s}{2}}v\right\rangle + \lambda\left\langle \mathcal{C}_0^{\frac{s-1}{2}}z, \mathcal{C}_0^{\frac{s-1}{2}}v\right\rangle = \left\langle \mathcal{C}_0^{-\frac{s}{2}}h, \mathcal{C}_0^{\frac{s}{2}}v\right\rangle,$$

where $z = \mathcal{C}_0^{-\frac{s}{2}}z' \in X^{1-s}$. Choosing $v = z \in X^{1-s}$, we get

$$\left\|\mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}\mathcal{C}_0^{\frac{s}{2}}z\right\|^2 + \lambda\left\|\mathcal{C}_0^{\frac{s-1}{2}}z\right\|^2 = \left\langle h, z\right\rangle.$$

By the Assumption 2.4(2) and the Cauchy-Schwarz inequality, we have

$$\left\|z\right\|_{\beta-2\ell-s}^2 + \lambda\left\|z\right\|_{1-s}^2 \leq c\left\|h\right\|_{-\eta_3}\left\|z\right\|_{\eta_3}.$$

We interpolate the norm of z appearing on the right hand side between the norms of z appearing on the left hand side, to get as in the proof of Proposition 5.1, for $c > 0$ independent of θ_3, λ and s

$$\|z\|_{\beta-2\ell-s} \le c\lambda^{-\frac{\theta_3}{2}} \|h\|_{-\eta_3} \quad \text{and} \quad \|z\|_{1-s} \le c\lambda^{-\frac{\theta_3+1}{2}} \|h\|_{-\eta_3}.$$

Replacing $z = \mathcal{C}_0^{-\frac{s}{2}} \mathcal{B}_\lambda^{-1} \mathcal{C}_0^{-\frac{s}{2}} h$ gives the first two rates.

For the last claim, note that we can always choose $\{\beta - 2\ell\} \vee \{s_0\} < s \le 1$, since $s_0 < 1$ and $\Delta > 0$. Using the first two estimates, for $\eta_{30} = (1 - \theta_{30})(\beta - 2\ell - s) + \theta_{30}(1 - s) = 0$, that is $\theta_{30} = \frac{2\ell-\beta+s}{\Delta} \in [0,1]$, we have that

$$\left\|\mathcal{C}_0^{-\frac{s}{2}} \mathcal{B}_\lambda^{-1} \mathcal{C}_0^{-\frac{s}{2}}\right\|_{\mathcal{L}(X^{\beta-2\ell-s}, \mathcal{X})} \le c\lambda^{-\frac{\theta_{30}}{2}}$$

and

$$\left\|\mathcal{C}_0^{-\frac{s}{2}} \mathcal{B}_\lambda^{-1} \mathcal{C}_0^{-\frac{s}{2}}\right\|_{\mathcal{L}(X^{1-s}, \mathcal{X})} \le c\lambda^{-\frac{\theta_{30}+1}{2}}.$$

Let $u \in \mathcal{X}$. Then, for any $t > 0$, we have the decomposition

$$u = \sum_{k=1}^{\infty} u_k \phi_k = \sum_{\lambda_k^{-1} \le t} u_k \phi_k + \sum_{\lambda_k^{-1} > t} u_k \phi_k =: \underline{u} + \overline{u},$$

where $\{\phi_k\}_{k=1}^{\infty}$ are the eigenfunctions of $\mathcal{C}_0$ and $u_k := \langle u, \phi_k \rangle$. Since $1 - s \ge 0$ and $\beta - 2\ell - s < 0$, we have

$$\left\|\mathcal{C}_0^{-\frac{s}{2}} \mathcal{B}_\lambda^{-1} \mathcal{C}_0^{-\frac{s}{2}} u\right\| \le \left\|\mathcal{C}_0^{-\frac{s}{2}} \mathcal{B}_\lambda^{-1} \mathcal{C}_0^{-\frac{s}{2}} \underline{u}\right\| + \left\|\mathcal{C}_0^{-\frac{s}{2}} \mathcal{B}_\lambda^{-1} \mathcal{C}_0^{-\frac{s}{2}} \overline{u}\right\|$$

$$\le c\lambda^{-\frac{\theta_{30}+1}{2}} \|\underline{u}\|_{1-s} + c\lambda^{-\frac{\theta_{30}}{2}} \|\overline{u}\|_{\beta-2\ell-s}$$

$$= c\lambda^{-\frac{\theta_{30}+1}{2}} \left(\sum_{\lambda_k^{-1} \le t} \lambda_k^{2s-2} u_k^2\right)^{-\frac{1}{2}} + c\lambda^{-\frac{\theta_{30}}{2}} \left(\sum_{\lambda_k^{-1} > t} \lambda_k^{2s+4\ell-2\beta} u_k^2\right)^{-\frac{1}{2}}$$

$$\le c\lambda^{-\frac{\theta_{30}+1}{2}} t^{1-s} \|u\| + c\lambda^{-\frac{\theta_{30}}{2}} t^{\beta-2\ell-s} \|u\|.$$

The first term on the right hand side is increasing in t, while the second is decreasing, so we can optimize by choosing $t = t(\lambda)$ making the two terms equal, that is $t = \lambda^{\frac{1}{2\Delta}}$, to obtain the claimed rate. $\qquad\square$

6. Posterior Contraction. In this section we employ the developments of the preceding sections to study the posterior consistency of the Bayesian solution to the inverse problem. That is, we consider a family of data sets $y^\dagger = y^\dagger(n)$ given by (1.10) and study the limiting behavior of the posterior measure $\mu_{\lambda,n}^{y^\dagger} = \mathcal{N}(m_\lambda^\dagger, \mathcal{C})$ as $n \to \infty$. Intuitively we would hope to recover a measure which concentrates near

the true solution $u^\dagger$ in this limit. Following the approach in [14], [9], [24] and [8], we quantify this idea as in (1.12). By the Markov inequality we have

$$\mathbb{E}^{y^\dagger} \mu_{\lambda,n}^{y^\dagger} \left\{ u : \left\| u - u^\dagger \right\| \geq M_n \varepsilon_n \right\} \leq \frac{1}{M_n^2 \varepsilon_n^2} \mathbb{E}^{y^\dagger} \int \left\| u - u^\dagger \right\|^2 \mu_{\lambda,n}^{y^\dagger}(du),$$

so that it suffices to show that

(6.1)
$$\mathbb{E}^{y^\dagger} \int \left\| u - u^\dagger \right\|^2 \mu_{\lambda,n}^{y^\dagger}(du) \leq c\varepsilon_n^2.$$

In addition to n^{-1}, there is a second small parameter in the problem, namely the regularization parameter, λ, and we will choose a relationship between n and λ in order to optimize the convergence rates ε_n. We will show that determination of optimal convergence rates follows directly from the operator norm bounds on $\mathcal{B}_\lambda^{-1}$ derived in the previous section, which concern only λ dependence; relating n to λ then follows as a trivial optimization. Thus, the λ dependence of the operator norm bounds in the previous section forms the heart of the posterior contraction analysis.

We now present our convergence results. In Theorem 6.1 and Corollary 6.4 we study the convergence of the posterior mean to the true solution in a range of norms, while in Theorem 6.2 and Corollary 6.5 we study the concentration of the posterior near the true solution as described in (1.12). The proofs of the two theorems are provided later in the current section.

THEOREM 6.1.   *Let* $u^\dagger \in X^1$. *Under the Assumptions 2.1 and 2.4, we have that, for the choice* $\tau = \tau(n) = n^{\frac{\theta_2 - \theta_1 - 1}{2(\theta_1 - \theta_2 + 2)}}$ *and for any* $\theta \in [0,1]$

$$\mathbb{E}^{y^\dagger} \left\| m_\lambda^\dagger - u^\dagger \right\|_\eta^2 \leq cn^{\frac{\theta + \theta_2 - 2}{\theta_1 - \theta_2 + 2}},$$

where $\eta = (1-\theta)(\beta - 2\ell) + \theta$. *The result holds for any* $\theta_1, \theta_2 \in [0,1]$, *chosen so that* $\mathbb{E}(\kappa^2) < \infty$, *for* $\kappa = \max \left\{ \left\| \xi \right\|_{2\beta - 2\ell - \eta_1}, \left\| u^\dagger \right\|_{2-\eta_2} \right\}$, *where* $\eta_i = (1-\theta_i)(\beta - 2\ell) + \theta_i$, $i = 1, 2$.

THEOREM 6.2.   *Let* $u^\dagger \in X^1$. *Under the Assumptions 2.1 and 2.4, we have that, for* $\tau = \tau(n) = n^{\frac{\theta_2 - \theta_1 - 1}{2(\theta_1 - \theta_2 + 2)}}$, *the convergence in (1.12) holds with*

$$\varepsilon_n = n^{\frac{\theta_0 + \theta_2 - 2}{2(\theta_1 - \theta_2 + 2)}}, \quad \theta_0 = \begin{cases} \frac{2\ell - \beta}{\Delta}, & \text{if } \beta - 2\ell \leq 0 \\ 0, & \text{otherwise.} \end{cases}$$

The result holds for any $\theta_1, \theta_2 \in [0,1]$, *chosen so that* $\mathbb{E}(\kappa^2) < \infty$, *for* $\kappa = \max \left\{ \left\| \xi \right\|_{2\beta - 2\ell - \eta_1}, \left\| u^\dagger \right\|_{2-\eta_2} \right\}$, *where* $\eta_i = (1-\theta_i)(\beta - 2\ell) + \theta_i$, $i = 1, 2$.

REMARK 6.3.   *i) In order to have convergence in the PDE method we need* $\mathbb{E}\left\| u^\dagger \right\|_{2-\eta_2}^2 < \infty$ *for a* $\theta_2 \leq 1$. *If we have the a-priori information that* $u^\dagger \in X^\gamma$, *then we need to have* $\gamma \geq 2 - \eta_2 = 1 + (1-\theta_2)\Delta$ *for some* $\theta_2 \in [0,1]$.

This means that the minimum requirement for convergence is $\gamma \geq 1$ which is compatible to our assumption $u^\dagger \in X^1$. On the other hand, since in order to have the optimal rate (which corresponds to choosing θ_2 as small as possible) we need to choose $\theta_2 = \frac{\Delta+1-\gamma}{\Delta}$, if $\gamma > 1 + \Delta$ then the right hand side is negative so we have to choose $\theta_2 = 0$, hence we cannot achieve the optimal rate. We say that the method saturates at $\gamma = 1 + \Delta$ which reflects the fact that the true solution has more regularity than the method allows us to exploit in order to obtain faster convergence rates.

ii) In order to have convergence we also need $\mathbb{E}\big\|\xi\big\|^2_{2\beta-2\ell-\eta_1} < \infty$ for a $\theta_1 \leq 1$. By Lemma *2.5(iii)*, it suffices to have $\theta_1 > \frac{s_0}{\Delta}$. This means that we need $\Delta > s_0$, which holds by the Assumption *2.4(1)*, in order to be able to choose $\theta_1 \leq 1$. On the other hand, since $\Delta > 0$ and $\sigma_0 \geq 1$, we have that $\frac{s_0}{\Delta} \geq 0$ thus we can always choose θ_1 in an optimal way, that is, we can always choose $\theta_1 = \frac{s_0+\varepsilon}{\Delta}$ where $\varepsilon > 0$ is arbitrarily small.

iii) If we want draws from μ_0 to be in X^γ then by Lemma *2.2(ii)* we need $\sigma_0 > \gamma$. Since the minimum requirement for the method to give convergence is $\gamma \geq 1$ while $\sigma_0 \leq 1$ this means that we can never have draws exactly matching the regularity of the prior. On the other hand if we want an undersmoothing prior (which according to [14] in the diagonal case gives asymptotic coverage equal to 1) we need $\sigma_0 \leq \gamma$, which we always have since $\gamma \geq 1$ and $\sigma_0 \leq 1$. This, as discussed in Section *1*, gives an explanation to the observation that in both of the above theorems we always have $\tau \to 0$ as $n \to \infty$.

iv) When $\beta - 2\ell > 0$, in Theorem *6.2* and in Corollary *6.5* below, we get suboptimal rates. The reason is that our analysis to obtain the error in the $\mathcal{X}$-norm is based on interpolating between the error in the $X^{\beta-2\ell}$-norm and the error in the X^1-norm. When $\beta - 2\ell > 0$, interpolation is not possible since the $\mathcal{X}$-norm is now weaker than the $X^{\beta-2\ell}$-norm. However, we can at least bound the error in the $\mathcal{X}$-norm by the error in the $X^{\beta-2\ell}$-norm, thus obtaining a suboptimal rate. Note, that the case $\beta - 2\ell > 0$ does not necessarily correspond to the well posed case: by Lemma *2.5* we can only guarantee that a draw from the noise distribution lives in X^ρ, $\rho < \beta - s_0$, while the range of $\mathcal{A}^{-1}$ is formally $X^{2\ell}$. Hence, in order to have a well posed problem we need $\beta - s_0 > 2\ell$, or equivalently $\Delta < \sigma_0$. This can happen despite our assumption $\Delta > 2s_0$, when $s_0 < 1/3$ and for appropriate choice of ℓ and β. In this case, regularization is unnecessary.

The following two corollaries are a direct consequence of the last remark:

COROLLARY 6.4.   *Assume $u^\dagger \in X^\gamma$, where $\gamma \geq 1$ and let $\eta = (1-\theta)(\beta - 2\ell) + \theta$, where $\theta \in [0, 1]$. Under the Assumptions 2.1 and 2.4, we have the following optimized rates of convergence, where $\varepsilon > 0$ is arbitrarily small:*

i) *if $\gamma \in (1, \Delta + 1]$, for $\tau = \tau(n) = n^{-\frac{\gamma-1+s_0+\varepsilon}{2(\Delta+\gamma-1+s_0+\varepsilon)}}$*

$$\mathbb{E}^{y^\dagger}\big\|m^\dagger_\lambda - u^\dagger\big\|^2_\eta \leq c n^{-\frac{\Delta+\gamma-1-\theta\Delta}{\Delta+\gamma-1+s_0+\varepsilon}};$$

ii) if $\gamma > \Delta + 1$, for $\tau = \tau(n) = n^{-\frac{\Delta+s_0+\varepsilon}{2(2\Delta+s_0+\varepsilon)}}$

$$\mathbb{E}^{y^\dagger}\left\|m_\lambda^\dagger - u^\dagger\right\|_\eta^2 \leq cn^{-\frac{(2-\theta)\Delta}{2\Delta+s_0+\varepsilon}};$$

iii) if $\gamma = 1$ and $\theta \in [0,1)$ for $\tau = \tau(n) = n^{-\frac{s_0+\varepsilon}{2(\Delta+s_0+\varepsilon)}}$

$$\mathbb{E}^{y^\dagger}\left\|m_\lambda^\dagger - u^\dagger\right\|_\eta^2 \leq cn^{-\frac{(1-\theta)\Delta}{\Delta+\gamma-1+s_0+\varepsilon}}.$$

If $\gamma = 1$ and $\theta = 1$ then the method does not give convergence;

COROLLARY 6.5.   *Assume $u^\dagger \in X^\gamma$, where $\gamma \geq 1$. Under the Assumptions 2.1 and 2.4, we have the following optimized rates for the convergence in (1.12), where $\varepsilon > 0$ is arbitrarily small:*

i) if $\gamma \in [1, \Delta+1]$ for $\tau = \tau(n) = n^{-\frac{\gamma-1+s_0+\varepsilon}{2(\Delta+\gamma-1+s_0+\varepsilon)}}$

$$\varepsilon_n = \begin{cases} n^{-\frac{\gamma}{2(\Delta+\gamma-1+s_0+\varepsilon)}}, & if\ \beta - 2\ell \leq 0 \\ n^{-\frac{\Delta+\gamma-1}{2(\Delta+\gamma-1+s_0\varepsilon)}}, & otherwise; \end{cases}$$

ii) if $\gamma > \Delta + 1$ for $\tau = \tau(n) = n^{-\frac{\Delta+s_0+\varepsilon}{2(2\Delta+s_0+\varepsilon)}}$

$$\varepsilon_n = \begin{cases} n^{-\frac{\Delta+1}{2(2\Delta+s_0+\varepsilon)}}, & if\ \beta - 2\ell \leq 0 \\ n^{-\frac{\Delta}{2\Delta+s_0+\varepsilon}}, & otherwise. \end{cases}$$

Note that, since the posterior is Gaussian, the left hand side in (6.1) is the Square Posterior Contraction

(6.2) $$SPC = \mathbb{E}^{y^\dagger}\left\|m_\lambda^\dagger - u^\dagger\right\|^2 + \mathrm{tr}(\mathcal{C}_{\lambda,\mathrm{n}}),$$

which is the sum of the mean integrated squared error (MISE) and the posterior spread. Let $u^\dagger \in X^1$. By Lemma 3.4, the relationship (1.10) between $u^\dagger$ and $y^\dagger$ and the equation (1.11) for $m_\lambda^\dagger$, we obtain

$$\mathcal{B}_\lambda m_\lambda^\dagger = \mathcal{A}^{-1}\mathcal{C}_1^{-1}y^\dagger = \mathcal{A}^{-1}\mathcal{C}^{-1}\mathcal{A}^{-1}u^\dagger + \frac{1}{\sqrt{n}}\mathcal{A}^{-1}\mathcal{C}^{-1}\xi$$

$$\text{and} \quad \mathcal{B}_\lambda u^\dagger = \mathcal{A}^{-1}\mathcal{C}^{-1}\mathcal{A}^{-1}u^\dagger + \lambda\mathcal{C}_0^{-1}u^\dagger,$$

where the equations hold in X^{-1}, since by a similar argument to the proof of Proposition 3.5 we have $m_\lambda^\dagger \in X^1$. By subtraction we get

$$\mathcal{B}_\lambda(m_\lambda^\dagger - u^\dagger) = \frac{1}{\sqrt{n}}\mathcal{A}^{-1}\mathcal{C}_1^{-1}\xi - \lambda\mathcal{C}_0^{-1}u^\dagger.$$

Therefore

(6.3) $$m_\lambda^\dagger - u^\dagger = \mathcal{B}_\lambda^{-1}\left(\frac{1}{\sqrt{n}}\mathcal{A}^{-1}\mathcal{C}_1^{-1}\xi - \lambda\mathcal{C}_0^{-1}u^\dagger\right),$$

as an equation in X^1. Using the fact that the noise has mean zero and the relation (1.6), equation (6.3) implies that we can split the square posterior contraction into three terms

$$(6.4) \qquad SPC = \left\| \lambda \mathcal{B}_\lambda^{-1} \mathcal{C}_0^{-1} u^\dagger \right\|^2 + \mathbb{E} \left\| \frac{1}{\sqrt{n}} \mathcal{B}_\lambda^{-1} \mathcal{A}^{-1} \mathcal{C}_1^{-1} \xi \right\|^2 + \frac{1}{n} \mathrm{tr}(\mathrm{B}_\lambda^{-1}),$$

provided the right hand side is finite. A consequence of the proof of Theorem 4.2 is that $\mathcal{B}_\lambda^{-1}$ is trace class. Note that for ζ a white noise, we have that

$$\mathrm{tr}(\mathcal{B}_\lambda^{-1}) = \mathbb{E} \left\| \mathcal{B}_\lambda^{-\frac{1}{2}} \zeta \right\|^2 = \mathbb{E} \langle \zeta, \mathcal{B}_\lambda^{-1} \zeta \rangle = \mathbb{E} \langle \mathcal{C}_0^{\frac{s}{2}} \zeta, \mathcal{C}_0^{-\frac{s}{2}} \mathcal{B}_\lambda^{-1} \mathcal{C}_0^{-\frac{s}{2}} \mathcal{C}_0^{\frac{s}{2}} \zeta \rangle$$

$$\leq \left\| \mathcal{C}_0^{-\frac{s}{2}} \mathcal{B}_\lambda^{-1} \mathcal{C}_0^{-\frac{s}{2}} \right\|_{\mathcal{L}(\mathcal{X})} \mathbb{E} \left\| \mathcal{C}_0^{\frac{s}{2}} \zeta \right\|^2,$$

which for $s > s_0$ since by Lemma 2.2 we have that $\mathbb{E} \left\| \mathcal{C}_0^{\frac{s}{2}} \zeta \right\|^2 < \infty$, provides the bound

$$(6.5) \qquad \mathrm{tr}(\mathcal{B}_\lambda^{-1}) \leq \mathrm{c} \left\| \mathcal{C}_0^{-\frac{s}{2}} \mathcal{B}_\lambda^{-1} \mathcal{C}_0^{-\frac{s}{2}} \right\|_{\mathcal{L}(\mathcal{X})},$$

where $c > 0$ is independent of λ. If q, r are chosen sufficiently large so that $\left\| \mathcal{C}_0^{-q} u^\dagger \right\| < \infty$ and $\mathbb{E} \left\| \mathcal{C}_0^r \xi \right\|^2 < \infty$ then we see that

(6.6)

$$SPC \leq c \left(\left\| \lambda \mathcal{B}_\lambda^{-1} \mathcal{C}_0^{q-1} \right\|_{\mathcal{L}(\mathcal{X})}^2 + \frac{1}{n} \left\| \mathcal{B}_\lambda^{-1} \mathcal{A}^{-1} \mathcal{C}_1^{-1} \mathcal{C}_0^{-r} \right\|_{\mathcal{L}(\mathcal{X})}^2 + \frac{1}{n} \left\| \mathcal{C}_0^{-\frac{s}{2}} \mathcal{B}_\lambda^{-1} \mathcal{C}_0^{-\frac{s}{2}} \right\|_{\mathcal{L}(\mathcal{X})} \right),$$

where $c > 0$ is independent of λ and n. Thus identifying ε_n in (1.12) can be achieved simply through properties of the inverse of $\mathcal{B}_\lambda$ and its parametric dependence on λ.

In the following, we are going to study convergence rates for the square posterior contraction, (6.4), which by the previous analysis will secure that

$$\mathbb{E}^{y^\dagger} \mu_{\lambda,n}^{y^\dagger} \left\{ u \colon \left\| u - u^\dagger \right\| \geq \varepsilon_n \right\} \to 0,$$

for $\varepsilon_n^2 \to 0$ at a rate almost as fast as the square posterior contraction. This suggests that the error is determined by the MISE and the trace of the posterior covariance, thus we optimize our analysis with respect to these two quantities. In [14] the situation where $\mathcal{C}_0, \mathcal{C}_1$ and $\mathcal{A}$ are diagonalizable in the same eigenbasis is studied, and it is shown that the third term in equation (6.4) is bounded by the second term in terms of their parametric dependence on λ. The same idea is used in the proof of Theorem 6.2.

We now provide the proofs of Theorem 6.1 and Theorem 6.2.

PROOF OF THEOREM 6.1. Since ξ has zero mean, we have by (6.3)

$$\mathbb{E} \left\| m_\lambda^\dagger - u^\dagger \right\|_{\beta-2\ell}^2 = \lambda^2 \left\| \mathcal{B}_\lambda^{-1} \mathcal{C}_0^{-1} u^\dagger \right\|_{\beta-2\ell}^2 + \frac{1}{n} \mathbb{E} \left\| \mathcal{B}_\lambda^{-1} \mathcal{A}^{-1} \mathcal{C}_1^{-1} \xi \right\|_{\beta-2\ell}^2$$

and
$$\mathbb{E}\big\|m_\lambda^\dagger - u^\dagger\big\|_1^2 = \lambda^2\big\|\mathcal{B}_\lambda^{-1}\mathcal{C}_0^{-1}u^\dagger\big\|_1^2 + \frac{1}{n}\mathbb{E}\big\|\mathcal{B}_\lambda^{-1}\mathcal{A}^{-1}\mathcal{C}_1^{-1}\xi\big\|_1^2.$$

Using Propositions 5.1 and 5.2, we get

$$\mathbb{E}\big\|m_\lambda^\dagger - u^\dagger\big\|_{\beta-2\ell}^2 \leq c\mathbb{E}(\kappa^2)(\lambda^{2-\theta_2} + \frac{1}{n}\lambda^{-\theta_1}) = c\mathbb{E}(\kappa^2)(n^{\theta_2-2}\tau^{2\theta_2-4} + n^{\theta_1-1}\tau^{2\theta_1})$$

and

$$\mathbb{E}\big\|m_\lambda^\dagger - u^\dagger\big\|_1^2 \leq c\mathbb{E}(\kappa^2)(\lambda^{1-\theta_2} + \frac{1}{n}\lambda^{-\theta_1-1}) = \frac{c\mathbb{E}(\kappa^2)}{\lambda}(n^{\theta_2-2}\tau^{2\theta_2-4} + n^{\theta_1-1}\tau^{2\theta_1}).$$

Since the common parenthesis term, consists of a decreasing and an increasing term in τ, we optimize the rate by choosing $\tau = \tau(n) = n^p$ such that the two terms become equal, that is, $p = \frac{\theta_2-\theta_1-1}{2(\theta_1-\theta_2+2)}$. We obtain,

$$\mathbb{E}\big\|m_\lambda^\dagger - u^\dagger\big\|_{\beta-2\ell}^2 \leq c\mathbb{E}(\kappa^2)n^{\frac{\theta_2-2}{\theta_1-\theta_2+2}} \quad \text{and} \quad \mathbb{E}\big\|m_\lambda^\dagger - u^\dagger\big\|_1^2 \leq c\mathbb{E}(\kappa^2)n^{\frac{\theta_2-1}{\theta_1-\theta_2+2}}.$$

By interpolating between the two last estimates we obtain the claimed rate.   □

PROOF OF THEOREM 6.2. Recall equation (6.4)

$$SPC = \big\|\lambda\mathcal{B}_\lambda^{-1}\mathcal{C}_0^{-1}u^\dagger\big\|^2 + \mathbb{E}\big\|\frac{1}{\sqrt{n}}\mathcal{B}_\lambda^{-1}\mathcal{A}^{-1}\mathcal{C}_1^{-1}\xi\big\|^2 + \frac{1}{n}\text{tr}(\mathrm{B}_\lambda^{-1}).$$

The idea is that the third term is always dominated by the second term. Combining equation (6.5) with Proposition 5.3, we have that

$$\frac{1}{n}\text{tr}(\mathcal{B}_\lambda^{-1}) \leq c\frac{1}{n}\lambda^{-\frac{2\ell-\beta+s}{\Delta}}, \ \forall s \in (\{\beta - 2\ell\} \vee \{s_0\}, 1].$$

i) Suppose $\beta - 2\ell \leq 0$, so that by interpolating between the rates provided by Proposition 5.1 and 5.2 we get for $\theta_0 = \frac{2\ell-\beta}{\Delta} \in [0,1]$

$$\mathbb{E}\big\|\frac{1}{\sqrt{n}}\mathcal{B}_\lambda^{-1}\mathcal{A}^{-1}\mathcal{C}_1^{-1}\xi\big\|^2 \leq \frac{1}{n}\mathbb{E}\big\|\xi\big\|_{2\beta-2\ell-\eta_1}^2\lambda^{-\theta_1-\theta_0}$$

and

$$\big\|\lambda\mathcal{B}_\lambda^{-1}\mathcal{C}_0^{-1}u^\dagger\big\|^2 \leq c\big\|u^\dagger\big\|_{2-\eta_2}^2\lambda^{2-\theta_2-\theta_0}.$$

Note that θ_1 is chosen so that $\mathbb{E}\big\|\xi\big\|_{2\beta-2\ell-\eta_1}^2 < \infty$, that is, by Lemma 2.5(iii), it suffices to have $\theta_1 > \frac{s_0}{\Delta}$. Noticing that by choosing s arbitrarily close to s_0, we can have $\frac{2\ell-\beta+s}{\Delta}$ arbitrarily close to $\frac{2\ell-\beta+\sigma_0}{\Delta}$, and since $\theta_1 + \theta_0 > \frac{2\ell-\beta+s_0}{\Delta}$, we deduce that the third term in equation (6.4) is always dominated by the second term. Combining, we have that

$$SPC \leq \frac{c\mathbb{E}(\kappa^2)}{\lambda^{\theta_0}}(\lambda^{2-\theta_2} + \frac{1}{n}\lambda^{-\theta_1}) = \frac{c\mathbb{E}(\kappa^2)}{\lambda^{\theta_0}}(n^{\theta_2-2}\tau^{2\theta_2-4} + n^{\theta_1-1}\tau^{2\theta_1}).$$

ii) Suppose $\beta - 2\ell > 0$. Using the Propositions 5.1 and 5.2 we have that

$$\left\|\lambda\mathcal{B}_\lambda^{-1}\mathcal{C}_0^{-1}u^\dagger\right\|^2 \leq c\left\|\lambda\mathcal{B}_\lambda^{-1}\mathcal{C}_0^{-1}u^\dagger\right\|_{\beta-2\ell}^2 \leq c\left\|u^\dagger\right\|_{2-\eta_2}^2\lambda^{2-\theta_2}$$

and

$$\mathbb{E}\left\|\frac{1}{\sqrt{n}}\mathcal{B}_\lambda^{-1}\mathcal{A}^{-1}\mathcal{C}_1^{-1}\xi\right\|^2 \leq c\mathbb{E}\left\|\frac{1}{\sqrt{n}}\mathcal{B}_\lambda^{-1}\mathcal{A}^{-1}\mathcal{C}_1^{-1}\xi\right\|_{\beta-2\ell}^2$$

$$\leq c\frac{1}{n}\mathbb{E}\left\|\xi\right\|_{2\beta-2\ell-\eta_1}^2\lambda^{-\theta_1},$$

where as before $\theta_1 > \frac{s_0}{\Delta}$. The third term in equation (6.4) is again dominated by the second term, since on the one hand $\theta_1 > \frac{s_0}{\Delta}$ and on the other hand, since $\beta - 2\ell > 0$, we can always choose $\{\beta - 2\ell\} \vee \{s_0\} < s \leq 1 \wedge \{s_0 + \beta - 2\ell\}$ to get $\frac{2\ell-\beta+s}{\Delta} \leq \frac{s_0}{\Delta}$. Combining the three estimates we have that

$$SPC \leq c\mathbb{E}(\kappa^2)(n^{\theta_2-2}\tau^{2\theta_2-4} + n^{\theta_1-1}\tau^{2\theta_1}).$$

In both cases, the common term in the parenthesis consists of a decreasing and an increasing term in τ, thus we can optimize by choosing $\tau = \tau(n) = n^p$ making the two terms equal, that is, $p = \frac{\theta_1-\theta_2+1}{2\theta_2-2\theta_1-4}$, to get the claimed rates. $\square$

7. Example. In this section we present a nontrivial example satisfying Assumptions 2.1 and 2.4.

Let $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$ be a bounded and open set. We define $\mathcal{A}_0 := -\mathbf{\Delta}$, where $\mathbf{\Delta}$ is the Dirichlet Laplacian which is the Friedrichs extension of the classical Laplacian defined on $C_0^2(\Omega)$, that is, $\mathcal{A}_0$ is a self-adjoint operator with a domain $\mathcal{D}(\mathcal{A}_0)$ dense in $L^2(\Omega)$ [15]. For $\partial\Omega$ sufficiently smooth we have $\mathcal{D}(\mathcal{A}_0) = H^2(\Omega) \cap H_0^1(\Omega)$. It is well known that $\mathcal{A}_0$ has a compact inverse and that it possesses an eigensystem $\{\rho_k^2, e_k\}_{k=1}^\infty$, where the eigenfunctions $\{e_k\}$ form a complete orthonormal basis of $L^2(\Omega)$ and the eigenvalues ρ_k^2 behave asymptotically like $k^{\frac{2}{d}}$ [1].

We study the Bayesian inversion of the operator $\mathcal{A} := \mathcal{A}_0 + \mathcal{M}_q$ where $\mathcal{M}_q : L^2(\Omega) \to L^2(\Omega)$ is the multiplication operator by a nonnegative function $q \in W^{2,\infty}(\Omega)$. Note that by the Hölder inequality the operator $\mathcal{M}_q$ is bounded. We assume that the observational noise is white, so that $\mathcal{C}_1 = I$, and we set the prior covariance operator to be $\mathcal{C}_0 = \mathcal{A}_0^{-2}$.

The operator $\mathcal{C}_0$ is trace class. Indeed, let $\lambda_k^2 = \rho_k^{-4}$ be its eigenvalues. Then they behave asymptotically like $k^{-\frac{4}{d}}$ and $\sum_{k=1}^\infty k^{-\frac{4}{d}} < \infty$ for $d < 4$. Furthermore, we have that $\sum_{k=1}^\infty \lambda_k^{2(1-\sigma)} \leq c\sum_{k=1}^\infty k^{-\frac{4(1-\sigma)}{d}} < \infty$, provided $\sigma < 1 - \frac{d}{4}$, that is, the Assumption 2.1 is satisfied with

$$\sigma_0 = \begin{cases} 3/4, & d = 1 \\ 1/2, & d = 2 \\ 1/4, & d = 3. \end{cases}$$

We define the Hilbert scale induced by $\mathcal{C}_0 = \mathcal{A}_0^{-2}$, that is, $(X^s)_{s\in\mathbb{R}}$, for $X^s := \overline{\mathcal{M}}^{\|\cdot\|_s}$, where

$$\mathcal{M} = \bigcap_{k=0}^\infty \mathcal{D}(\mathcal{A}_0^{2k}), \ \langle u, v\rangle_s := \langle\mathcal{A}_0^s u, \mathcal{A}_0^s v\rangle \quad \text{and} \quad \|u\|_s := \|\mathcal{A}_0^s u\|.$$

Observe, $X^0 = \mathcal{X} = L^2(\Omega)$.

Our aim is to show that $\mathcal{C}_1 \simeq \mathcal{C}_0^\beta$ and $\mathcal{A}^{-1} \simeq \mathcal{C}_0^\ell$, where $\beta = 0$ and $\ell = \frac{1}{2}$, in the sense of the Assumptions 2.4. We have $\Delta = 2\ell - \beta + 1 = 2$. Since for $d = 1, 2, 3$ we have $0 < s_0 < 1$, the Assumption 2.4(1) is satisfied. Moreover, note that since $\mathcal{C}_1 = I$ the Assumptions 2.4(3) and (4) are trivially satisfied.

We now show that Assumptions 2.4 (2), (5), (6) are also satisfied. In this example the three assumptions have the form

2. $\left\|(\mathcal{A}_0 + \mathcal{M}_q)^{-1}u\right\| \asymp \left\|\mathcal{A}_0^{-1}u\right\|$, $\forall u \in X^{-1}$;
5. $\left\|\mathcal{A}_0^s(\mathcal{A}_0 + \mathcal{M}_q)^{-1}u\right\| \le c_3\left\|\mathcal{A}_0^{s-1}u\right\|$, $\forall u \in X^{s-1}$, $\forall s \in (s_0, 1]$;
6. $\left\|\mathcal{A}_0^{-\eta}(\mathcal{A}_0 + \mathcal{M}_q)^{-1}u\right\| \le c_4\left\|\mathcal{A}_0^{-\eta-1}u\right\|$, $\forall u \in X^{-\eta-1}$, $\forall \eta \in [-1, 1]$.

Observe that Assumption (5) is implied by Assumption (6).

LEMMA 7.1. *The operator* $(\mathcal{A}_0 + \mathcal{M}_q)^{-1}\mathcal{A}_0$ *is bounded when considered as an operator (i)* $\mathcal{X} \to \mathcal{X}$, *(ii)* $X^{-1} \to X^{-1}$ *and (iii)* $X^1 \to X^1$.

PROOF.

i) $(\mathcal{A}_0 + \mathcal{M}_q)^{-1}\mathcal{A}_0 = (I + \mathcal{A}_0^{-1}\mathcal{M}_q)^{-1}$, where $\mathcal{A}_0^{-1}\colon \mathcal{X} \to \mathcal{X}$ is compact and $\mathcal{M}_q\colon \mathcal{X} \to \mathcal{X}$ is bounded thus $K := \mathcal{A}_0^{-1}\mathcal{M}_q\colon \mathcal{X} \to \mathcal{X}$ is compact. Hence, by the Fredholm Alternative [12, §27, Theorem 9], it suffices to show that -1 is not an eigenvalue of K. Indeed, if there exists $u \in \mathcal{X}$ such that $\mathcal{A}_0^{-1}\mathcal{M}_q u = -u$, then $\mathcal{M}_q u = -\mathcal{A}_0 u$, therefore u satisfies $(\mathcal{A}_0 + \mathcal{M}_q)u = 0$. Since $\mathcal{A}_0 + \mathcal{M}_q$ is positive-definite we have that $u = 0$, thus -1 is not an eigenvalue of K.

ii) The claim is equivalent to the inequality

$$\left\|\mathcal{A}_0^{-1}(\mathcal{A}_0 + \mathcal{M}_q)^{-1}u\right\| \le c\left\|\mathcal{A}_0^{-2}u\right\|, \ \forall u \in X^{-2}.$$

Put $v = (\mathcal{A}_0 + \mathcal{M}_q)^{-1}u$ and note that we want to estimate $w = \mathcal{A}_0^{-1}v$, where v satisfies $\mathcal{A}_0 v + \mathcal{M}_q v = u$, or multiplying by $\mathcal{A}_0^{-2}$, $w + \mathcal{A}_0^{-2}\mathcal{M}_q\mathcal{A}_0 w = \mathcal{A}_0^{-2}u$. The operator $K = \mathcal{A}_0^{-2}\mathcal{M}_q\mathcal{A}_0$ is compact in $\mathcal{X}$. Indeed, since $\mathcal{A}_0^{-1}$ is compact, it suffices to show that $\mathcal{A}_0^{-1}\mathcal{M}_q\mathcal{A}_0$ is bounded in $\mathcal{X}$. By duality for $u \in \mathcal{X}$ we have

$$\left\|\mathcal{A}_0^{-1}\mathcal{M}_q\mathcal{A}_0 u\right\| = \sup_{\|\phi\|=1}\left\langle \mathcal{A}_0^{-1}\mathcal{M}_q\mathcal{A}_0 u, \phi\right\rangle = \sup_{\|\phi\|=1}\left\langle u, \mathcal{A}_0\mathcal{M}_q\mathcal{A}_0^{-1}\phi\right\rangle$$

$$\le \|u\| \sup_{\|\phi\|=1}\left\|\mathcal{A}_0\mathcal{M}_q\mathcal{A}_0^{-1}\phi\right\| \le c\|u\|\|q\|_{W^{2,\infty}(\Omega)},$$

since

$$\left\|\mathcal{A}_0\mathcal{M}_q\mathcal{A}_0^{-1}\phi\right\| = \left\|\mathbf{\Delta}\mathcal{M}_q\mathcal{A}_0^{-1}\phi\right\| = \left\|(\mathbf{\Delta}q)\mathcal{A}_0^{-1}\phi + 2(\nabla q)(\nabla\mathcal{A}_0^{-1}\phi) + q\mathbf{\Delta}\mathcal{A}_0^{-1}\phi\right\|$$

$$\le \|q\|_{W^{2,\infty}(\Omega)}(\left\|\mathcal{A}_0^{-1}\phi\right\| + \left\|\nabla\mathcal{A}_0^{-1}\phi\right\| + \|\phi\|) \le c\|q\|_{W^{2,\infty}(\Omega)}\|\phi\|.$$

Note that -1 cannot be an eigenvalue of K since in that case we would have $\mathcal{A}_0^{-2}\mathcal{M}_q\mathcal{A}_0 u = -u$ for some $u \in \mathcal{X} \setminus \{0\}$ thus $\mathcal{M}_q\mathcal{A}_0 u = -\mathcal{A}_0^2 u$ and setting

$\psi = \mathcal{A}_0 u$ we would get $\mathcal{A}_0 \psi + \mathcal{M}_q \psi = 0$ thus since $\mathcal{A}_0 + \mathcal{M}_q$ is positive-definite $\psi = 0$ and since $\mathcal{A}_0$ is positive-definite $u = 0$. By the Fredholm Alternative we have that $(I + K)^{-1}$ is bounded in $\mathcal{X}$, hence $\|w\| \leq \|(I+K)^{-1}\|_{\mathcal{L}(\mathcal{X})} \|\mathcal{A}_0^{-2} u\|$ or equivalently

$$\|\mathcal{A}_0^{-1}(\mathcal{A}_0 + \mathcal{M}_q)^{-1} u\| \leq c\|\mathcal{A}_0^{-2} u\|, \ \forall u \in X^{-2}.$$

iii) The claim is equivalent to the inequality

$$\|\mathcal{A}_0(\mathcal{A}_0 + \mathcal{M}_q)^{-1} u\| \leq c\|u\|, \ \forall u \in \mathcal{X}.$$

Put $v = (\mathcal{A}_0 + \mathcal{M}_q)^{-1} u$ and note that we want to estimate $w = \mathcal{A}_0 v$, where v satisfies $\mathcal{A}_0 v + \mathcal{M}_q v = u$ or equivalently $w + \mathcal{M}_q \mathcal{A}_0^{-1} w = u$. The operator $K = \mathcal{M}_q \mathcal{A}_0^{-1}$ is compact in $\mathcal{X}$, since $\mathcal{A}_0^{-1}$ is compact and $\mathcal{M}_q$ is bounded in $\mathcal{X}$. Note that -1 cannot be an eigenvalue of K since if there exists a $u \in \mathcal{X} \setminus \{0\}$ such that $\mathcal{M}_q \mathcal{A}_0^{-1} u = -u$, then setting $\mathcal{A}_0^{-1} u = \psi$ we have that $\mathcal{M}_q \psi = -\mathcal{A}_0 \psi$ thus since $\mathcal{A}_0 + \mathcal{M}_q$ is positive-definite we have $\psi = 0$ and since $\mathcal{A}_0^{-1}$ is positive-definite we have $u = 0$. By the Fredholm Alternative we conclude that $(I + K)^{-1}$ is bounded in $\mathcal{X}$, hence $\|w\| \leq \|(I+K)^{-1}\|_{\mathcal{L}(\mathcal{X})} \|u\|$ or equivalently that

$$\|\mathcal{A}_0(\mathcal{A}_0 + \mathcal{M}_q)^{-1} u\| \leq c\|u\|, \ \forall u \in \mathcal{X}.$$

$\square$

By direct application of [17, Theorem 4.36 and Theorem 1.18] we have the following interpolation result:

PROPOSITION 7.2. *The couple (X^0, X^1) is an interpolation couple and for every $\theta \in [0, 1]$ we have $(X^0, X^1)_{\theta, 2} = X^\theta$. Furthermore, the couple (X^0, X^{-1}) is also an interpolation couple and for every $\theta \in [0, 1]$ we have $(X^0, X^{-1})_{\theta, 2} = X^{-\theta}$.*

PROPOSITION 7.3. *The Assumptions 2.4 are satisfied for this example.*

PROOF. We need to show that the Assumptions 2.4(2) and (6) hold.

i) We first prove Assumption 2.4(2). By (i) of the last lemma we have

$$\|(\mathcal{A}_0 + \mathcal{M}_q)^{-1} u\| \leq c\|\mathcal{A}_0^{-1} u\|, \ \forall u \in X^{-1}.$$

We need to show that $\|(\mathcal{A}_0 + \mathcal{M}_q)^{-1} u\| \geq c\|\mathcal{A}_0^{-1} u\|, \ \forall u \in X^{-1}$ which is equivalent to $\|\mathcal{A}_0^{-1}(\mathcal{A}_0 + \mathcal{M}_q) u\| \leq c\|u\|, \ \forall u \in \mathcal{X}$. Indeed,

$$\|\mathcal{A}_0^{-1}(\mathcal{A}_0 + \mathcal{M}_q) u\| \leq \|u\| + \|\mathcal{A}_0^{-1}(qu)\| \leq (1 + \|q\|_\infty) \|u\|,$$

since $\mathcal{A}_0^{-1}$ is bounded and by the Hölder inequality.

ii) For the proof of Assumption 2.4(6) it suffices to show that $L := (\mathcal{A}_0 + \mathcal{M}_q)^{-1} \mathcal{A}_0 \in \mathcal{L}(X^\eta), \ \forall \eta \in [-1, 1]$. By Lemma 7.1 we have that $L \in \mathcal{L}(X^0) \cap \mathcal{L}(X^1)$ and $L \in \mathcal{L}(X^0) \cap \mathcal{L}(X^{-1})$, thus by [17][Theorem 1.6] and the last proposition, we get the result.

$\square$

We can now apply Corollary 6.4 and Corollary 6.5 to get the following convergence result.

THEOREM 7.4. *Let* $u^\dagger \in X^\gamma, \gamma \geq 1$. *Then, for* $\tau = \tau(n) = n^{\frac{4-d-4\gamma+\varepsilon}{8\gamma+8+2d+\varepsilon}}$, *the convergence in (1.12) holds with* $\varepsilon_n = n^{-e}$, *where*

$$e = \begin{cases} \frac{2\gamma}{4+d+4\gamma+\varepsilon}, & \text{if } \gamma < 3 \\ \frac{6}{16+d+\varepsilon}, & \text{if } \gamma \geq 3, \end{cases}$$

for $\varepsilon > 0$ *arbitrarily small and where* $d = 1,2,3$, *is the dimension. Furthermore, for* $t \in [-1,1]$, *for the same choice of* τ, *we have* $\mathbb{E}\|m_\lambda^\dagger - u^\dagger\|_t^2 \leq cn^{-h}$, *where*

$$h = \begin{cases} \frac{4\gamma-4t}{4+d+4\gamma+\varepsilon}, & \text{if } \gamma < 3 \\ \frac{12-4t}{16+d+\varepsilon}, & \text{if } \gamma \geq 3. \end{cases}$$

8. The Diagonal Case. In the case where $\mathcal{C}_0, \mathcal{C}_1$ and $\mathcal{A}$, are all diagonalizable in the same eigenbasis our assumptions are trivially satisfied, provided $\Delta > 2s_0$. In [14], sharp convergence rates are obtained for the convergence in (1.12), in the case where the three relevant operators are simultaneously diagonalizable and have spectra that decay algebraically; the authors only consider the case $\mathcal{C}_1 = I$ since in this diagonal setting the colored noise problem can be reduced to the white noise one. The rates in [14] agree with the minimax rates provided the scaling of the prior is optimally chosen, [4]. In Figure 1 (cf. Section 1) we have in green the rates of convergence predicted by Corollary 6.5 and in blue the sharp convergence rates from [14], plotted against the regularity of the true solution, $u^\dagger \in X^\gamma$, in the case where $\beta = \ell = \frac{1}{2}$ and $\mathcal{C}_0$ has eigenvalues that decay like k^{-2}. In this case $s_0 = \frac{1}{2}$ and $\Delta = \frac{3}{2}$, so that $\Delta > 2s_0$.

As explained in Remark 6.3, the minimum regularity for our method to work is $\gamma = 1$ and our rates saturate at $\gamma = 1 + \Delta$, that is, in this example at $\gamma = 2.5$. We note that for $\gamma \in [1, 2.5]$ our rates agree, up to $\varepsilon > 0$ arbitrarily small, with the sharp rates obtained in [14], for $\gamma > 2.5$ our rates are suboptimal and for $\gamma < 1$ the method fails. In [14], the convergence rates are obtained for $\gamma > 0$ and the saturation point is at $\gamma = 2\Delta$, that is, in this example at $\gamma = 3$. In general the PDE method can saturate earlier (if $2\ell - \beta < 0$), at the same time (if $2\ell - \beta = 0$), or later (if $2\ell - \beta > 0$) compared to the diagonal method presented in [14]. However, the case $2\ell - \beta > 0$ in which our method saturates later, is also the case in which our rates are suboptimal, as explained in Remark 6.3(iv).

The discrepancies can be explained by the fact that in Proposition 5.2, the choice of θ_2 which determines both the minimum requirement on the regularity of $u^\dagger$ and the saturation point, is the same for both of the operator norm bounds. This means that on the one hand to get convergence of the term $\|\lambda \mathcal{B}_\lambda^{-1} \mathcal{C}_0^{-1} u^\dagger\|$ in equation (6.4) in the proof of Theorem 6.2, we require conditions which secure the convergence in the stronger X^1-norm and on the other hand the saturation rate for this term

is the same as the saturation rate in the weaker $X^{\beta-2\ell}$-norm. For example, when $\beta - 2\ell = 0$ the saturation rate in the PDE method is the rate of the $\mathcal{X}$-norm hence the agreement of the saturation point with the rates in [14]. In particular, we have agreement of the saturation rate when $\beta = \ell = 0$, which corresponds to the problem where we directly observe the unknown function polluted by white noise (termed the *white noise model*).

9. Conclusions. We have presented a new method of identifying the posterior distribution in a conjugate Gaussian Bayesian linear inverse problem setting (Section 4). We used this identification to examine the posterior consistency of the Bayesian approach in a frequentist sense (Section 6). We provided convergence rates for the convergence of the expectation of the mean error in a range of norms (Theorem 6.1, Corollary 6.4). We also provided convergence rates for the square posterior contraction (Theorem 6.2, Corollary 6.5). Our methodology assumed a relation between the prior covariance, the noise covariance and the forward operator, expressed in the form of norm equivalence relations (Assumptions 2.4). We considered Gaussian noise which can be white. In order for our methods to work we required a certain degree of ill-posedness compared to the regularity of the prior (Assumption 2.4(1)) and for the convergence rates to be valid a certain degree of regularity of the true solution. In the case where the three involved operators are all diagonalizable in the same eigenbasis, when the problem is sufficiently ill-posed and for a range of values of γ, the parameter expressing the regularity of the true solution, our rates agree (up to $\varepsilon > 0$ arbitrarily small) with the sharp (minimax) convergence rates obtained in [14] (Section 8).

The methodology presented in this paper is extended to drift estimation for diffusion processes in [19]. Future research includes the extension to an abstract setting which includes both the present paper and [19] as special cases. Other possible directions are the consideration of nonlinear inverse problems, the use of non-Gaussian priors and/or noise and the extension of the credibility analysis presented in [14] to a more general setting.

REFERENCES

[1] ATTOUCH, H., BUTTAZZO, G. and MICHAILLE, G. (2006). *Variational analysis in Sobolev and BV spaces. MPS/SIAM Series on Optimization* **6**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Applications to PDEs and optimization. MR2192832

[2] BANKS, H. T. and KUNISCH, K. (1989). *Estimation techniques for distributed parameter systems. Systems & Control: Foundations & Applications* **1**. Birkhäuser Boston Inc., Boston, MA. MR1045629

[3] BOGACHEV, V. I. (1998). *Gaussian measures. Mathematical Surveys and Monographs* **62**. American Mathematical Society, Providence, RI. MR1642391

[4] CAVALIER, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems* **24** 034004, 19. . MR2421941

[5] DA PRATO, G. (2006). *An introduction to infinite-dimensional analysis. Universitext.* Springer-Verlag, Berlin. Revised and extended from the 2001 original by Da Prato. MR2244975

[6] DIACONIS, P. and FREEDMAN, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14** 1–67. With a discussion and a rejoinder by the authors. . MR829555

[7] ENGL, H. W., HANKE, M. and NEUBAUER, A. (1996). *Regularization of inverse problems. Mathematics and its Applications* **375**. Kluwer Academic Publishers Group, Dordrecht. MR1408680

[8] FLORENS, J.-P. and SIMONI, A. (2010). Regularizing Priors for Linear Inverse Problems. *Preprint.*

[9] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. . MR1790007

[10] GINÉ, E. and NICKL, R. (2011). Rates of contraction for posterior distributions in L^r- metrics, $1 \leq r \leq \infty$. *Ann. Statist.* **39** 2883-2911.

[11] HAIRER, M., STUART, A. M. and VOSS, J. (2007). Analysis of SPDEs arising in path sampling. II. The nonlinear case. *Ann. Appl. Probab.* **17** 1657–1706. . MR2358638

[12] HELMBERG, G. (1969). *Introduction to spectral theory in Hilbert space. North-Holland Series in Applied Mathematics and Mechanics, Vol. 6.* North-Holland Publishing Co., Amsterdam. MR0243367

[13] KALLENBERG, O. (1997). *Foundations of modern probability. Probability and its Applications (New York).* Springer-Verlag, New York. MR1464694

[14] KNAPIK, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.* **39** 2626-2657.

[15] LAX, P. D. (2002). *Functional analysis. Pure and Applied Mathematics (New York).* Wiley-Interscience [John Wiley & Sons], New York. MR1892228

[16] LEHTINEN, M. S., PÄIVÄRINTA, L. and SOMERSALO, E. (1989). Linear inverse problems for generalised random variables. *Inverse Problems* **5** 599–612. MR1009041

[17] LUNARDI, A. (2009). *Interpolation theory*, second ed. *Appunti. Scuola Normale Superiore di Pisa (Nuova Serie). [Lecture Notes. Scuola Normale Superiore di Pisa (New Series)].* Edizioni della Normale, Pisa. MR2523200

[18] MANDELBAUM, A. (1984). Linear estimators and measurable linear transformations on a Hilbert space. *Z. Wahrsch. Verw. Gebiete* **65** 385–397. . MR731228

[19] POKERN, Y., STUART, A. M. and VAN ZANTEN, J. H. (2012). Posterior Consistency via Precision Operators for Bayesian Nonparametric Drift Estimation in SDEs. Accepted Stoch. Proc. Appl. http://arxiv.org/abs/1202.0976.

[20] ROBINSON, J. C. (2001). *Infinite-dimensional dynamical systems. Cambridge Texts in Applied Mathematics.* Cambridge University Press, Cambridge. An introduction to dissipative parabolic PDEs and the theory of global attractors. . MR1881888 (2003f:37001a)

[21] ROUSSEAU, J. (2010). Rates of convergence for the posterior distributions of mixtures of betas and adaptive nonparametric estimation of the density. *Ann. Statist.* **38** 146–180. . MR2589319

[22] SHEN, X. and WASSERMAN, L. (2001). Rates of convergence of posterior distributions. *Ann. Statist.* **29** 687–714. . MR1865337

[23] STUART, A. M. (2010). Inverse problems: a Bayesian perspective. *Acta Numer.* **19** 451–559. . MR2652785

[24] VAN DER VAART, A. W. and H. VAN ZANTEN, J. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* **36** 1435–1463. . MR2418663

[25] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2007). Bayesian inference with rescaled Gaussian process priors. *Electron. J. Stat.* **1** 433–448 (electronic). . MR2357712

MATHEMATICS INSTITUTE
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL, UK
E-MAIL: s.agapiou@warwick.ac.uk
        A.M.Stuart@warwick.ac.uk

MATHEMATICAL SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
AND UNIVERSITY OF GOTHENBURG
SE-412 96 GOTHENBURG, SWEDEN
E-MAIL: stig@chalmers.se